ABSTRACT

Title of dissertation: LOCAL INFORMATION LANDSCAPES: THEORY, MEASURES, AND EVIDENCE

Myeong Lee
Doctor of Philosophy, 2019

Dissertation directed by: Professor Brian S. Butler
College of Information Studies

To understand issues about information accessibility within communities, research studies have examined human, social, and technical factors by taking a socio-technical view. While this view provides a profound understanding of how people seek, use, and access information, this approach tends to overlook the impact of the larger structures of information landscapes that constantly shape peoples access to information. When it comes to local community settings where local information is embedded in diverse material entities such as urban places and technical infrastructures, the effect of information landscapes should be taken into account in addition to particular strategies for solving information-seeking issues.

However, characterizing the information landscape of a local community at the community level is a non-trivial problem due to diverse contexts, users, and their interactions with each other. One way to conceptualize local information landscapes in a way that copes with the complexity of the interplay between information, contexts, and human factors is to focus on the materiality of information. By focusing

on the material aspects of information, it becomes possible to understand how local information is provided to social entities and infrastructures and how it exists, forming structures at the community level. Through an extensive literature review, this paper develops a theory of local information landscapes (LIL Theory) to better conceptualize the community-level, material structure of local information. Specifically, the LIL theory adapts a concept of the virtual as an ontological view of the interplay between technical infrastructures, spaces, and people as a basis for assessing and explaining community-level structures of local information. By complementing existing theories such as information worlds and information grounds, this work provides a new perspective on how information deserts manifest as a material precondition of information inequality.

Using this framework, an empirical study was conducted to examine the explicit effects of information deserts on other community characteristics. Specifically, the study aims to provide an initial assessment of LIL theory by examining how the fragmentation of local information, a form of information deserts, is related to important community characteristics such as socio-economic inequality, deprivation, and community engagement. Building upon previous work in sociology and political science, this study shows that the fragmentation of local information (1) is shaped by socio-economic deprivation/inequality that is confounded with ethnoracial heterogeneity, (2) the fragmentation of local information is highly correlated to people's community gatherings, (3) the fragmentation of local information moderates the effects of socio-economic inequality on cultural activity diversity, and (4) the fragmentation of local information mediates the relationship between socio-economic

inequality and community engagement. By making use of three local event datasets over 20 months in 14 U.S. cities (about two million records) and over 3 months in 28 U.S. cities (about 620K records), respectively, this study develops computational frameworks to operationalize information deserts in a scalable way.

This dissertation provides a theorization of community-level information inequality and computational models that support the quantitative examination of it. Further theorizations of the conceptual constructs and methodological improvements on measurements will benefit information policy-makers, local information system designers, and researchers who study local communities with conceptual models, vocabularies, and assessment frameworks.

LOCAL INFORMATION LANDSCAPES: THEORY, MEASURES, AND EVIDENCE

by

Myeong Lee

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2019

Advisory Committee:
Professor Brian S. Butler, Chair
Professor Paul Jaeger
Assistant Professor Vanessa Frias-Martinez
Assistant Professor Grant McKenzie
Associate Professor Hiroyuki Iseki, Dean's Representative

Acknowledgments

My journey throughout the doctoral program at the UMD iSchool has been full of joy. First of all, I thank God for making all these things happen. So many people helped me complete this dissertation.

Dr. Brian Butler has been a great mentor, a supervisor, and an advisor since my first year of the MIM program in 2012. His support and guidance for the past seven years have been an invaluable resource and immense encouragement for me. I thank his patience and affection even when I was not making progress.

I am grateful to the other dissertation committee members for their help and support. Dr. Vanessa Frias-Martinez and Dr. Grant McKenzie encouraged and guided me to explore diverse computational techniques and social-good spirit. Dr. Paul Jaeger and Dr. Hiro Iseki provided insightful comments on the theories and methods in my dissertation and fine-tuned this dissertation. It would have been impossible to finish this without the committee members' support.

I also thank Dr. Susan Winter, Dr. Ping Wang, Dr. Richard Marciano, and Dr. Joel Chan for their collaboration and support throughout my Ph.D. journey. Thanks to them, I was able to explore a broader spectrum of iSchool research, beyond the scope of my dissertation, through co-authoring interesting papers and grant proposals. Other iSchool faculty members have encouraged and inspired me tremendously as well. Dr. Wayne Lutters, Dr. Eun Kyoung Choe, Dr. Ken Heger, Dr. Amanda Lazar, Dr. Beth St. Jean, Andy Fellows, and Dr. Michael Kurtz have been amazing teachers and exemplary faculty members.

I would not have gone through this program without my friends and colleagues. I appreciate Jonathan and Christie for their friendship and hang-outs year long. I really enjoyed being caffeinated, having meals together, and discussing research and life affairs. I am also thankful to Chiyoung, Marina, Jinyoung, Joohee, Xu, Leyla, Andrew, David, Lingzi, Yuting, Priya, and Mary for their check-ins, discussions, friendship, and thoughtful responses to my silly ideas. Edel, Yuheng, Shiyun, and Lauren have been excellent team members for DCIC projects in the last couple of years. I really appreciate their friendship and engagement.

I learned a lot through the fellowship programs and summer institutes. It was an unforgettable experience for me to work with brilliant researchers, practitioners, and colleagues through these programs. Particularly, I would like to thank the Data Science for Social Good programs at Georgia Tech (Dr. Chris Le Dantec and team members) and at the University of Washington's eScience Institute (Dr. Afra Mashhadi, Dr. Ariel Rokem, and team members) where I was able to learn about whole new sets of knowledge and approaches. I also thank the Summer Institute of Computational Social Science at NYU (Dr. Adaner Usmani and Barum Park), the Center for Open Data Enterprise (Joel Gurin, Audrey Ariss, and Laura Manley), and CSST Summer Institutes (Dr. Rosta Farzan, Dr. Claudia López, and Dr. Bryan Semaan) who set my theoretical and philosophical grounds.

I am grateful to my colleagues and friends outside my institutions who have greatly supported me and helped finish this dissertation. Megan Weng has been a great supervisor for my software development work at UMD. Dr. Daniele Quercia's work emboldened a novice iSchooler to explore computational social sciences along with his warm and welcoming friendship. Minsu Park, an inspiring friend, helped me greatly with understanding CSS culture and recommended the key literature. Also, Sangseok, Javier, and Dr. John Harlow encouraged me and provided useful comments on my research.

During my time in the U.S., my old friends and mentors have always been there for me. Dr. Seongsoo Hong, Hyejun, and Jonghun have supported me from the beginning of my graduate life. I send my gratitude to them. I thank Jinwoo, Jaewook, Seunghee, Kyungseok, Sangbum, Yongjin, Jaehyun, Chang-Goo, and Soo-Hyun for their support and friendship since college. Dongwoo, Soohwan, Kyemyung, Deok Gun, and Eunsoo have been great friends who shared doctoral life stories as international students. I am also grateful to friends from St. Andrew Kim Church.

I want to say thank you to Kyounghwa and Yongmin for coming to my defense from distance, and special thanks to Dooheon for his continued friendship since elementary school and for coming to my defense all the way from New York. He has always inspired me with interesting ideas.

I send my love to my family who have always been proud of me, and to my grandparents who are in heaven. I am proud to be your son, grandson, and brother. Finally, I send my utmost love and gratitude to my fiancé, Hyein, who has always listened to my research and daily affairs with empathy, encouraged me to go through hardships, and inspired me to think about new ideas. All this work wouldn't have been possible without her endless support and love throughout my graduate studies.

On a hot summer day in July, 2019

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Motivation

Information access issues such as how libraries and other information institutions enable people to find desired information effectively have long been of interest to information science researchers [Burnett et al., 2008, Culnan, 1985]. Disparities in information access have been sometimes characterized as the *digital divide* or *information inequality* due to people's high dependency on technology in finding information and its high impact on their socio-economic status [Van Dijk, 2005, Norris, 2001, Schiller, 2013]. Because many factors are associated with these problems, information access has been studied from two perspectives: a human-centered view and a technology-focused view. A human-centered view is based on the understanding of individual- or group-level characteristics such as physical impairment [Malu and Findlater, 2015], economic status [Smith and Hanisch, 2015], generational factors [Russell and Young, 2015], contextualized experience [Lloyd, 2005] and education level [Boer, 2015]. A technology-focused view is one that focuses on system or material factors such as the information aggregation/filtering technique [Kavanaugh et al., 2014], information visibility [Struppek, 2006], and recommendation system performance [Lee and Brusilovsky, 2017]. Of course, these issues are not solely about humans or technology; rather, a tension exists between the two approaches, placing each study somewhere on a *socio-technical* spectrum [Langefors, 1978].

The socio-technical view provides valuable insight into the factors that affect people's information access in various contexts and situations. At the same time, studies taking the socio-technical approach assume that (1) information is socially-constructed, (2) contexts where information is created and used, i.e., *information grounds* [Fisher and Naumer, 2006], shape the dynamics of people in accessing information along with individual and technological factors, (3) information is somewhere in the virtual or physical world, and (4) the distribution and structure of available information is determined by socio-technical strategies and practices such as efficient data organization in a particular context. These assumptions lead to the idea that understanding contexts and users, designing a proper information system in a contextualized way, and understanding users and technology in a holistic way are the main concerns when addressing information access problems. They provide an effective basis for identifying important aspects of individual- and organizational-level information access issues such as information seeking, using, and sharing.

In a local community or urban environment, human-centered and technology-focused approaches can still provide a useful basis for understanding different types of information access problems. However, studies that take a socio-technical view are more likely to focus on the information within particular systems that are used by specific groups of people. As a result, it is possible for information science researchers and system designers to overlook the larger systems of information in a community and their implications for both individuals and the community. In other words, socio-technical approaches risk underestimating the effects of the complex structures of available information and local information sources that constantly shape individuals' information behavior and access, regardless of new information systems. If a researcher studies a digital literacy problem of residents in a city by examining their use of particular online calendars and websites for seeking local event information, for example, this researcher may have to target only a part of the event information available in the entire city, maybe 20% at most, according to a study by López et al. [2014]. It would be possible to understand residents information behavior and barriers in using the target information sources, but the researcher may face difficulties in explaining their holistic capacity that stems from other existing infrastructures and information sources. Even if people have access to local event information through social media, there could be other important information that is available only on physical bulletin boards in a neighborhood, or that is not available at all in formal online and offline locations, but only disseminated through a limited number of people.

This suggests that understanding the structure of local information at the community level can provide useful perspectives on individuals' information access and behavior. However, conceptualizing community-level information landscapes is not straightforward because there are many different users and contexts interacting with each other in a community, which create diverse boundaries of systems through which information is made available. Therefore, the motivation of this dissertation is at understanding the structure and features of local information at the community level and their relationships to other community characteristics by reconceptualizing them in the way that reduces the complexity of information users and contexts. Based on the theoretical and empirical gaps identified from the literature review, this dissertation develops frameworks and methods to operationalize the components of local information landscapes.

1.2 Information Access and Accessibility of Information

Access to and *accessibility* of information are conceptualized differently in the literature. Information access is defined as "the presence of a robust system through which information is made available" [Burnett et al., 2008, Jaeger and Burnett, 2005], while information accessibility is conceptualized by focusing more on the organization and presentation of information itself [Fidel and Green, 2004]. Information often becomes accessible through implementing proper access structures such as formats and services [Hill, 2013]. The notion of information accessibility would be meaningful when a particular audience is presumed because "accessibility" can be defined as a counterbalance to potential users' characteristics in the conceptual space of actors and artifacts [Fidel and Green, 2004]. For example, information access of engineers who seek field-specific information is facilitated through their awareness of information sources and physical proximity to them [Fidel and Green, 2004]. As such, information accessibility is often maximized by increasing sources' visibility and their physical proximity to particular groups of users.

When the term "information accessibility" is used without considering a particular group of people, it conceptualizes the characteristic of information itself with presumably a general audience in mind, thus “universal accessibility” or “universal usability” [Hill, 2013, Shneiderman, 2000]. This concept might be useful in many places, but could be limited in making solutions generalizable, because each context and user group has distinctive characteristics and different ways of interpreting information and interacting it. Due to this reason, information accessibility is usually considered along with the dimensions of information access in understanding people’s contextualized information behavior and use. While some articles use these terms interchangeably depending on the focus of research, this dissertation uses the terms precisely: “information access” is used to describe information users’ systematic connections with information sources and “information accessibility” emphasizes more on the characteristics of information and information sources themselves.

These two different concepts pose interesting questions when it comes to community-level information landscapes. On the one hand, the dimensions of information access, such as people’s physical, intellectual, and social access to information [Jaeger and Burnett, 2005], inform key variables that are related to the structure and features of local information at the community level. Even if a piece of information itself is accessible from a data quality and access structure perspective, it is possible that some populations might not have access to it because of their specific traits. These characteristics that shape information access can be better identified with through the conceptual frameworks. On the other hand, the accessibility of local information could be re-conceptualized when it comes to the community level. Studies on particular contexts and users often focus on information quality, data formats, and support services. However, these accessibility-related factors might provide only partial understanding of information accessibility, when the focus of research is on whether people have access to the entire information that is available in a community. Conceptually, this dimension of information accessibility is highly dependent on the structure and distribution of information rather than its source-level organization and quality.

Theorizing local information landscapes, in these regards, can provide a new perspective on both information access and information accessibility by (1) providing new theoretical constructs and (2) extending the meaning of it, respectively, at the community level.

1.3 An Example: Broadband Accessibility

The term, *community-level information access*, might remind many people of broadband accessibility. Because of people's high dependency on the internet in finding information today, telecommunication- and public-policy researchers have focused on this issue for a long time [Grubesic, 2012, Gabel and Kwan, 2001]. A longitudinal study on OECD countries, for example, showed that about a 10 percent increase in broadband penetration in a country positively affects annual per capita growth by approximately 0.9 to 1.5 percent [Czernich et al., 2011]. This kind of a large-scale study has influenced government agencies in developing fine-resolution visualization systems that demonstrate broadband accessibility issues in the United States (Figure 1.1). Such maps have been used as baseline evidence in implementing

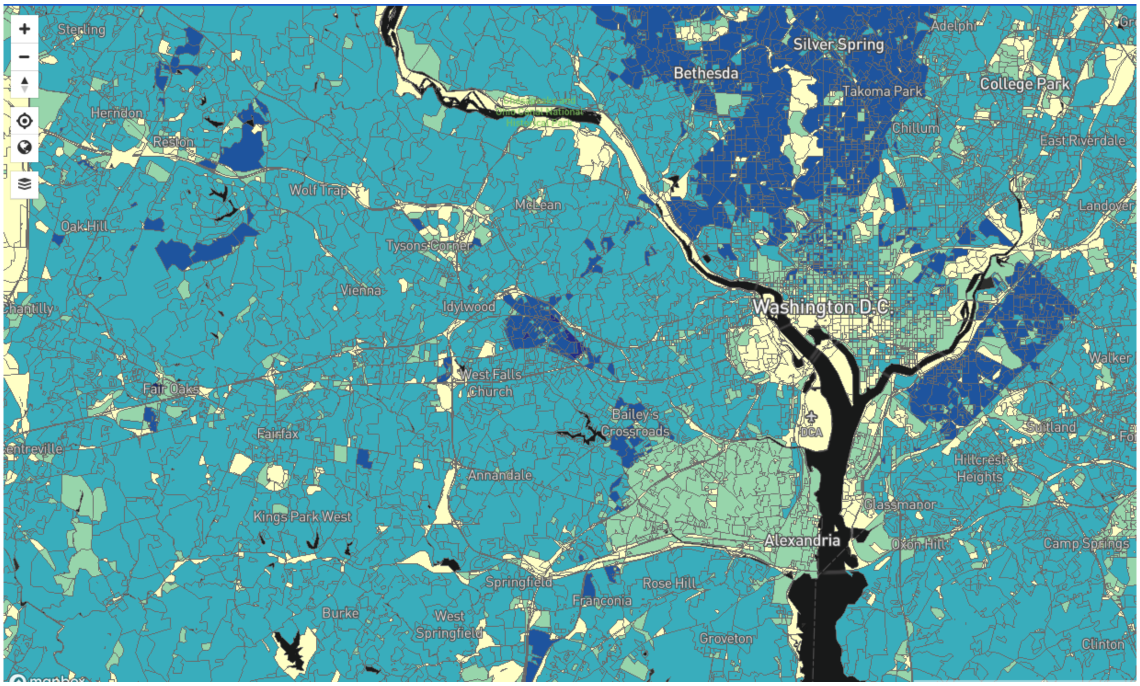

Figure 1.1: An example of broadband accessibility maps around Washington D.C., United States.[Federal Communication Commissions, 2019]

information policies about local information infrastructures such as the Broadband Data Services Improvement Act.[1]

While it is true that broadband penetration is one important factor in local communities that give rise to the socio-economic status of its residents, this factor explains only one aspect of local information access issues. Even before the internet was created, people connected with each other using diverse information infrastructures such as word-of-mouth, public libraries, local newspapers, and local bulletin boards. Although the emergence of the internet might have changed the dynamics and forms of local information infrastructures significantly, focusing only on technological infrastructures can risk underestimating the importance of existing human and institutional infrastructures in understanding people's information access and

[1]Original copy accessed as of April 2019: `https://www.govinfo.gov/content/pkg/PLAW-110publ385/pdf/PLAW-110publ385.pdf`

behaviors [Bowker and Star, 1998, Borgman, 2003].

Acknowledging this potential limitation, communications scholars conceptualized different levels of material access to technology, namely, the first-level, second-level, and third-level digital divide [van Deursen and van Dijk, 2019, Van Deursen and Helsper, 2015]. The first-level digital divide denotes people's disparities in basic infrastructure access, such as broadband access; the second-level digital divide conceptualizes inequality that stems from interface-level technology use such as computers, mobile devices, and tablets; and the third-level digital divide refers to the inequality in the outcomes of internet use or benefits derived from it [van Deursen and van Dijk, 2019]. These conceptualizations help overcome the limitation of focusing solely on broadband accessibility to some degree by articulating different levels of technology use and their impact on information access.

From an information user perspective, however, a conceptual gap still exists; that is, the availability of local information could be different from the availability of infrastructures and technological devices. Even if local bulletin boards are installed or high-speed broadband networks are available, it is possible that organizations and people do not provide sufficient information through those systems. This suggests that the information layer in a community needs to be understood separately from the entities and infrastructures that provide access to and embed information. In other words, if using van Deursen and van Dijk [2019]'s terminologies, there might be a conceptual and procedural gap that relates to *information itself* between the first-/second-level digital divide (i.e., disparities in technology access) and the third-level digital divide (i.e., inequality in internet use outcomes).

The two different layers, that is, infrastructures/technology and information that are available from them, might be closely related to each other, but their effects and people's behavioral patterns in using them could be completely different. If there is a more general, holistic way for describing local information and local infrastructures that embed it at the community level, information policies and related research studies will be able to take these diverse aspects of infrastructural and information issues into account, which will ultimately help create more equitable and democratic information landscapes in local communities. The first step towards this direction begins with their conceptualizations: how can we understand local information and infrastructures at the community level?

1.4 Research Goals and Organization

This dissertation aims to conceptualize the structure and features of local information from a community perspective and to empirically show how these community-level forces shape community outcomes such as community engagement. Through a literature review, this dissertation first theorizes local information landscapes that consider not only human and technological factors that affect individuals' information access, but also the impact of the complexity and diversity of the overall local information that exists in forms that range from an ad-hoc information system to a large organization. In order to do so, a conceptual model is constructed in a material way by viewing overall local information in a community as an entity that has structures and features and exists on top of diverse infrastructures and spaces.

In other words, by focusing on the materiality of information, the complexities that stem from diverse users and contexts are shifted from the foreground to the background [Dourish, 2017]. An analogy for the materiality of information can be drawn from discussions on the materiality of technology. Technology can be understood as material by itself before it is understood as a socially-constructed entity where its roles depend on social contexts and the people who use it [Leonardi et al., 2012]. Similarly, scholars conceptualized the materiality of information because information is presented in material forms such as monitor displays, flyers on bulletin boards, and magnetic tapes, which provide different capabilities for information users [Dourish, 2017, Dourish and Mazmanian, 2011]. The materiality of information is a pre- or necessary condition of the social construction of information. Also, it allows us to understand the forms and structure of local information at the community level by mitigating its entanglements with diverse contexts and users.

Based on this conceptualization process, the structure and features of local information, or *local information landscapes*, are suggested as important elements that may affect people's information behavior and access and mediate the relationships between community components/characteristics (e.g., geography, technical infrastructure, local organizations) and community outcomes (e.g., economic wellness, psychological satisfaction, civic participation). This material view provides opportunities to identify otherwise overlooked aspects of local information; specifically it leads us to think about (1) the creation of information landscapes that is closely related to *the provision of information*, (2) the relationships between local information landscapes and other community features such as civic engagement and economic well-being, (3) the conceptual differences between information behavior and provision, (4) the understanding of local spaces, people, and technical infrastructures as *material entities* that embed and provide local information, and (5) how the material pre-conditions of information inequality in a local community, i.e., *information deserts*, look and how they can be conceptualized and assessed in a systematic way.

The concept of *information deserts* makes it possible to identify diverse forms of structures that can give rise to information inequality in a local community. From a theoretical perspective, researchers who study city- or local community-related subjects can benefit from the proposed theory, because it provides a new lens and vocabularies to think about the relationship between information provision, community features, community outcomes, and information access. Practically, LIL theory and the operationalizations of information deserts can help local information system designers and policy makers understand information-related problems in a holistic way in their planning stage by asking questions such as, How would a new information system change the overall information access in a community?

Following the theorizing process of local information landscapes, computational approaches to process data and quantitative analysis/results are presented. In particular, an empirical study that makes use of LIL theory is designed by building upon sociology and political science research. This study aims to understand how the characteristics of local information landscapes affect or are affected by other community characteristics (particularly, community engagement). For the empirical study, computational methods are used to disambiguate pieces of information across different websites and to quantify the fragmentation of local information landscapes, a dimension of information deserts.

After reporting the analysis results, the dissertation discusses the implications of this new approach, methodological challenges, and future research. Through the empirical studies and computational modeling, this dissertation sheds light on the relationships between community-level, information-driven factors, and other community characteristics that were hardly captured or explained by previous studies, contributing to both the theories of community engagement and information behavior/access.

Chapter 2: Theoretical Framework: A Theory of Local Information Landscapes

2.1 Theoretical Models for Information Access, Behavior, and Context

To understand information access at the community level, it is necessary to review existing theories and models to see if they are feasible for describing community-level dynamics. Some theories and models that explain information behavior, access, and contexts are relevant for describing the relationships between local information-related factors and other community factors. They include, but are not limited to, models developed in information science, human-computer interaction (HCI), and communication. Among many other theories about information behavior and access, this section focuses on five models that exhibit a reasonable variability in their coverage of theoretical components because they provide an initial understanding about current streams of local information research and help identify potential risks and challenges in the research programs.

The *theory of information worlds*, which integrates Chatmans small world concept and Habermas' notion of the public sphere, explains multi-level influences that affect peoples information access and behaviors [Jaeger and Burnett, 2010, Chatman, 2000, Habermas, 1991]. Social boundaries and influences around individuals that vary from small world influences, such as a family and friends, to life-world structures, such as institutions and technology, shape peoples information behaviors and activities. According to this theory, small world and life-world influences are not two separate constructs, but two endpoints of a social influence spectrum, while still interacting with each other [Jaeger and Burnett, 2010]. Also, the theory of information worlds explains different kinds of information access issues, i.e., physical, intellectual, and social access [Burnett and Jaeger, 2008], that are affected by various environmental and contextual influences from the user perspective. For example, viewed from the information worlds lens, the *CiVicinity* project [Hoffman et al., 2012], a web-based local information portal that combines and filters local news and events from multiple information sources, can be interpreted in a boundary perspective; that is, the new technology not only enhances individuals' information access and use, but also expands technological and institutional boundaries by consolidating diverse information sources, i.e., an element of normative information behavior [Burnett and Jaeger, 2008]. While this theoretical lens allows us to understand information technology as part of the social and environmental influences that shape individuals information behavior, it marginally conceptualizes how information is created, accessed, and acquired by an individual, which is an essential part in describing the community-level structure of information.

Contextualizing and focusing more on people's interactions with the environment, Lloyd [2010] suggested the concept of the *information literacy landscapes* to describe the settings in which information literacy is understood and manifested. According to her, information landscapes is a socially-constructed, intersubjectively-created space that stems from people's interactions, collective experiences, and information practices, which in turn shape shared knowledge among them [Lloyd, 2010]. Given this conceptual space, information literacy is understood as people's activities and skills that are required to process and deal with the shared knowledge and situated information [Lloyd, 2006, 2010]. At a workplace, for example, an information landscape is formed through the interactions between an individual's literacy and other employees' actions; in an educational environment, it is created mostly based on students' literacy and textual information in books [Lloyd, 2006]. In these examples, information is a socially-constructed entity and a result or medium of the interaction between the information user and provider. Accordingly, the information landscapes of individuals change rapidly with regard to contexts and their contextualized experiences, shaping different landscapes for one another. Also, this conceptualization leads us to understand information inequality as a phenomenon that is closely related to people's literacy and their interactions with the diverse contexts in which they are embedded. It provides an insightful understanding of the interactions between information providers and receivers. However, at the same time, analysis becomes intractable when it comes to the community level due to the complex entanglements between diverse users, providers, and contexts.

Compared to the theories of information worlds and information literacy landscapes, the model of *information grounds* highlights the situational aspect, i.e., *grand contexts* where information is generated [Fisher et al., 2004, Fisher and Naumer, 2006]. In an empirical study of information grounds [Pettigrew, 1999], community health nurses' information behavior was examined during their services for elderly people. This study found that nurses' information giving behavior, i.e., *information provision*, was affected by four elements of information grounds that collectively created a grand context: environmental factors (e.g., type of building, weather), activities (e.g., treatment processes), factors associated with nurses (e.g., knowledge of local resources and senior's situations), and factors associated with patients. These elements partially overlap with those from the theory of information worlds (e.g., the layout of a building can be explained by the "explicit boundary" idea in information worlds). Savolainen [2009] compared these two models regarding spatial and social factors and concluded that they complemented each other by emphasizing different aspects. Spatial and social factors are normative constraints in a small world, while they are opportunities that enable serendipitous information seeking and sharing from the information grounds perspective. As such, the model of information grounds systematically explains factors and contexts that shape the information provision process by individuals. However, this model is still limited in conceptualizing diverse forms of local information in a holistic manner, thus providing rare implications on the community-level structure of local information.

A study from HCI also conceptualized the creation and flow of local information in neighborhoods as *data-in-place*, as a result of implementing location-based technology [Taylor et al., 2015]. Unlike information grounds, local data was intentionally created by installing physical devices (e.g., PosterVote) to develop the notion of data technology in a local community as an ecosystem of devices and services. Specifically, they characterized local information as a thing that matters by interacting with residents, shapes contours by being transferred, reflects spatial/temporal/social structures, and forms small worlds where each has its own internal logic [Chatman, 2000]. By elaborating on the characteristics of data in a material form in a neighborhood (as opposed to an abstract form of it in the cloud servers), this model effectively rationalizes how data interacts with people and places, and thus how data technology can be designed in ways that enhance community engagement. While limited in including other forms of data such as social media data in the model and in explaining people's information seeking behavior due to the intentional creation of data, it combines the concept of small worlds with materialized data to better describe the impact of local technologies on people's lives.

Some studies in the field of communication also suggested a useful model for describing local information systems. As the communication research field has mainly viewed information technology as media for enhancing communications, technical infrastructures in a local community were seen as communication channels and conceptualized as *communication infrastructures* [Kim and Ball-Rokeach, 2006, Ball-Rokeach et al., 2001]. Since this approach is useful in interpreting and characterizing communications among local residents, it can be beneficial in explaining people's interactions that are a basis for civic engagement. Taking this view, the CiVicinity example can be understood as one of many communication infrastructures that facilitates information exchange among people in a neighborhood [Hoffman et al., 2012]. However, the focus of this model is limited to the flow and dissemination of information, i.e., communication, rather than forms and availability of information in a static and scalable manner. As a result, the dynamics and effects of local information itself are not well characterized by the communication infrastructure view.

Together, these theoretical models explain different aspects of individuals' information seeking, sharing, and using behaviors in community settings, as well as the impact of local information on residents' dynamics. Viewed at the community level, based on the coverage of these models, it is possible to construct a conceptual diagram that summarizes the relationships between community factors that affect information behavior/access/provision and community outcomes that are affected by people's information behavior/access/provision (Figure 2.1). Each theory or model is mapped to corresponding arrows that show causalities or effects explained by it. Also, information behavior and provision are not distinctive concepts in this model because information provision is one of the information sharing and seeking behaviors of an individual, so they are attached together in the diagram [Williamson, 1998, Pettigrew, 1999].

However, this preliminary model for local information research provides limited understanding for local information when it comes to the level of community, as it presents two major challenges: *an ontological challenge* where none of the theoretical models in Figure 2.1 conceptualizes how provided information exists in a local community and *a causality challenge* where it is hard to explain how information provision affects individuals' information behavior and how provided information itself is related to other community factors.

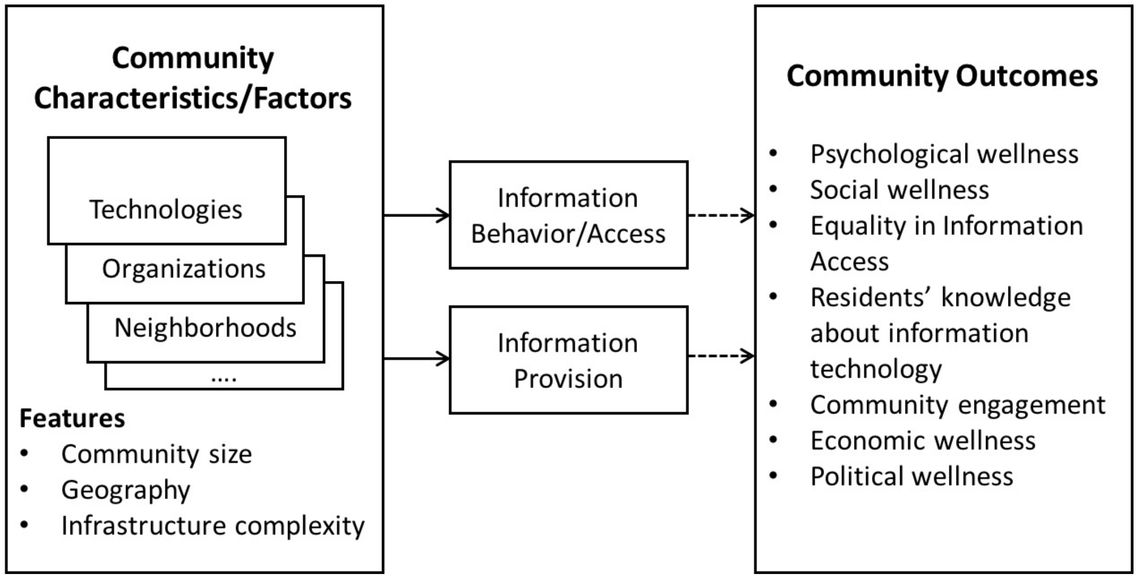

Figure 2.1: A preliminary model for the relationships between information behavior/access, local community factors, and outcomes based on the existing models and theories.

To understand the complexity and diversity of local information, its sources, and their interactions with each other, it is necessary to have a new model that addresses and articulates the structure and features of information available in a local community. This new conceptualization needs to be capable of (1) explaining different kinds of information available in a community in an abstract way, (2) providing an understanding of how complex relationships between diverse entities that contain information manifest, and (3) suggesting ways to describe different kinds of local information sources at the community level (rather than at the individual- or organization-level).

2.2 *The Virtual*: An Ontological View of Local Information Landscapes

The ontological challenge of the preliminary model (2.1) can be resolved through a new model that explains a community-level, material structure of local information. Among many other models and frameworks, the notion of *the virtual* is adapted to conceptualize how local information looks at the community level [De Souza e Silva and Sutko, 2011].

De Souza e Silva and Sutko [2011] conceptualized the interplay between people, space, technology, and information to understand the location-based technologies that people use in an urban environment based on the philosophy of *the virtual*. Unlike other communication theories, this model focuses less on technology as a communication medium, but rather recognizes location-based technologies as entities in the space of the virtual where one component *limits or extends* the capability of another among *people, technology, space*, and *information*.[1] The characterization of the interplay between components of the virtual, that is "limiting or extending the capability," aligns well with the concept of materiality, because materials (e.g., a computer) come with *capacity*, *ability*, or *affordances* before they are used in a meaningful way [Leonardi and Barley, 2008]. While not particularly developed as a material model of location-based technologies, the concept of the virtual can be adapted by incorporating the discussion of materialities of information [Dourish,

[1] In the conceptual space of the virtual, distinguishing between the virtual and physical spaces is less meaningful. They only limit or extend the capability of another between different entities.

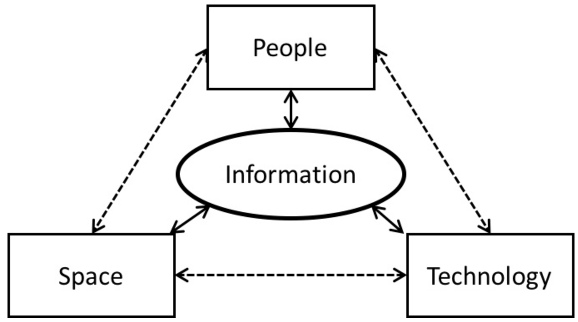

Figure 2.2: An adapted model of the virtual. Information can be embedded in either people, space, or technology in a material form, while each component extends or limits the capability of another.

2017].

Because people, spaces, and technologies are material entities that contain local information and, at the same time, provide representations of information, local information at the community level can be understood as being embedded in any of these entities while each component extends or limits the capacity of another (Figure 2.2). It explains a material state of how local information exists in a community and connotes the pre-condition of socially-constructed information (i.e., the condition of information before it becomes meaningful to people). Also, the adjusted model of the virtual denotes a material instantiation of *context*, because a combination of its components (e.g., nurses as people and buildings as spaces) is a necessary condition for a particular context before it is realized as socially-constructed elements (e.g., nurses as service providers and buildings as medical places[2]).

The embeddedness of information in material entities can be supported by

[2]The concepts of place and space follow the distinction made by Harrison and Dourish [1996], where space denotes the three-dimensional, physical structures while place includes cultural and social meanings attached to it.

previous studies about information behavior and access. The model of information grounds explains how information is created and provided to people. In the example of nurses in a community-based clinic [Pettigrew, 1999], information is given to elderly people by the influences of the environment (i.e., medical place), activity, and individual factors. From a material perspective, this sharing behavior is a provision of information to a social system. Meanwhile, other examples such as international students' wandering around behavior [Oh et al., 2014] and the fragmentation of local information sources [López et al., 2014] indicate that local information can be provided directly to physical spaces and technical infrastructures, respectively. These studies together, with the adapted model of the virtual, suggest that the process of local information provision can be described in any of three ways:

- Provided to a technological infrastructure (e.g., posting a story on social media)
- Provided to a physical space (e.g., posting a flyer on a bulletin board)
- Provided to a social system (e.g., notifying students about homework in a class)

In addition to the information provision process, the components of the virtual also have features such as scale, complexity, and persistence; it can be about space (small/homogeneous vs. large/heterogeneous), technology (small/simple vs. large/complex), and people (small group vs. complex social system). Also, the material characteristics of a component determine the persistence/transience of local

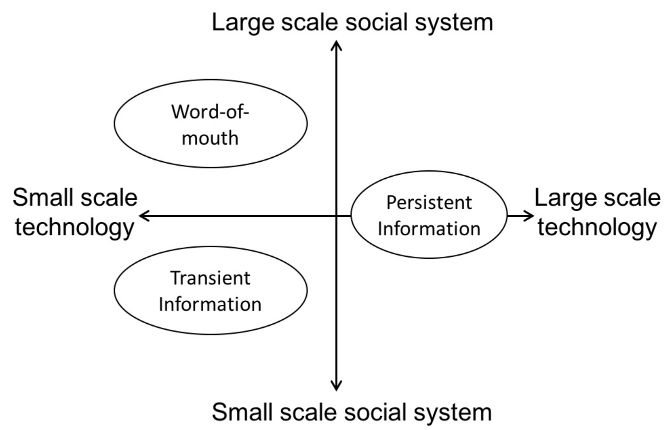

Figure 2.3: An example of the relationship between the scale of a component and information persistence.

information (e.g., if a small group of people forgot about particular information, information is transient; or if data a computer was gone due to a power outage, information is transient). An example of the relationship between the scale of a component and the characteristic of information is shown in Figure 2.3.

The scale, complexity, and persistence of each component, in turn, create a multi-dimensional model that better explains the material aspects of local information at the community level. In other words, the interplay between the components of the virtual and their material characteristics comprise a local information landscape. We define this adjusted concept of the virtual as *the model of local information landscapes* (*LIL model*). Based on this model, it is possible to refine the preliminary model of local community research (Figure 2.1). The new theoretical framework, i.e., *the LIL framework*, is presented in Figure 2.4. By inserting the model of local information landscapes in the preliminary model, it is possible to flesh out how information is provided to the community, how it exists, and the roles it

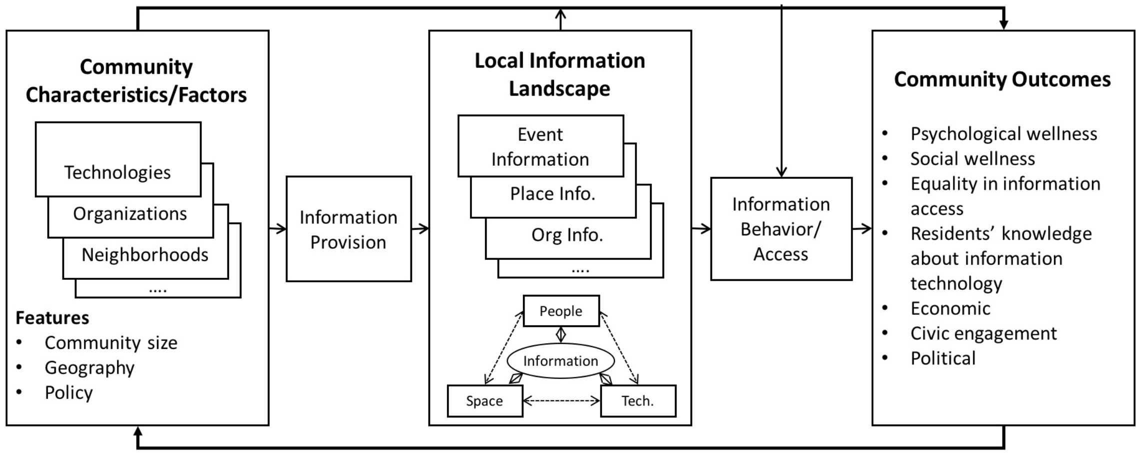

Figure 2.4: The LIL Framework: the key elements of the local information landscapes.

plays in shaping individuals' information behavior and other community outcomes. The LIL framework not only resolves the ontological challenge of the preliminary model, but also makes distinctions between information provision and behavior by viewing information in a material way.

The LIL framework allows us to focus on the structural effects of the technical and material systems in shaping other socio-technical dynamics of how people use, seek, and share information. Also, it implies that information availability in a community depends not only on individual or organizational activities to provide information to a community's spaces, digital fora, and/or social groups, but also how the interplay between the components of the virtual shapes their capabilities to embed and provide local information.

2.3 Previous Work on Local Communities and Local Information

The LIL framework integrates diverse studies about local communities, partially resolving the causality challenge of the preliminary model. To provide an initial assessment of the proposed framework's face validity, it is necessary to synthesize a broader spectrum of studies that deals with local communities and information. Throughout several academic disciplines, studies on this spectrum can be roughly categorized into three themes: (1) deployment of information and communication technologies (ICTs) to solve community problems, (2) correlation and causation between community features, LIL features, and community outcomes, and (3) examination of people's information behavior in local community settings.

2.3.1 Deployment of ICTs

Many studies from HCI and community informatics (CI) have focused on the psychological and human factors of residents and other community outcomes as dependent variables, and have designed and deployed technology as a means to solve local community problems. A representative work in this paradigm is Carroll's notions that characterize an important community outcome [Carroll, 2014]. He conceptualized local community factors with three facets: (1) community identity (attachment), (2) participation and awareness (engagement), and (3) social support network, which as a whole represents people's connectedness to the community. Having these kinds of community outcomes as main objects, community informatics and HCI researchers studied how various ICTs affected community outcomes such as people's community engagement and sense of connectedness to the community [Pasek et al., 2009, López and Farzan, 2015, Humphreys, 2010, Han et al., 2014, Le Dantec et al., 2015].

This trend has been predominant in the field of information and communication technologies for development (ICT4D), a sub-field of HCI and CI. Many studies implemented ICTs in local communities of developing countries and examined how technology influenced community outcomes. Examples include, but are not limited to, how increasing internet access through a community information center affected rural people's lives in Bangladesh [Hossain, 2015], the extent to which mobile phones eliminated non-commercial farmers' poverty [Smith and Hanisch, 2015], and how a technology literacy program sustained and enhanced local culture, beliefs, and traditions [Abd Aziz et al., 2015].

In this stream of studies, issues about people's information access, intermediary processes in achieving study goals, are not explicitly discussed and are often assumed as being resolved through the deployment of technological solutions. The presumption behind this might be that community outcomes such as socio-economic level, inequality, or civic engagement can be enhanced by either facilitating people's exchange of information or increasing their access to information that exists on the internet (without this assumption, for example, a technology influence on civic engagement is hardly explained). From the LIL perspective, these studies can be mapped onto the arrow between community features (i.e., technology) and outcomes (e.g., civic engagement) in the LIL framework (Figure 2.5 (a)).

Other studies in CI did not directly focus on community outcomes, but instead

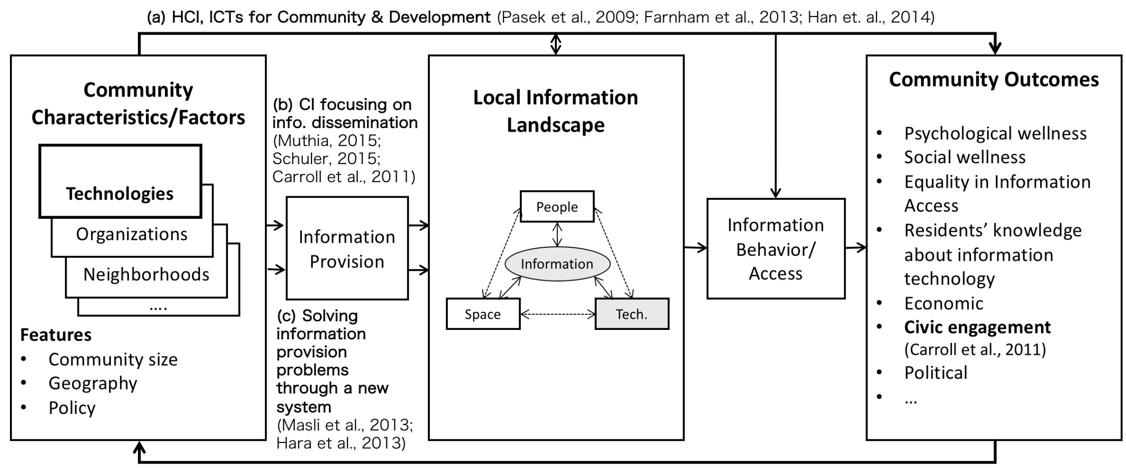

Figure 2.5: Mapping HCI and CI studies onto the LIL framework.

studied information access, provision, and dissemination. For example, Muthiah [2015] studied how a voice message service enhanced farmers' information dissemination in India and tried to understand if information dissemination actually helped reduce the poverty level in a community. Also, Schuler [2015] explored how academia could make community ICTs more relevant in the realm of civic intelligence. These research studies provide a methodological approach to generating information in a local community and imply that community-wide strategies are necessary for facilitating information provision. These types of studies can be mapped onto the arrow between community characteristics and information provision (Figure 2.5 (b)).

Design studies that focus on crowdsourcing platforms for local communities are another stream of research in the category of technology deployment. Volunteered geographic information (VGI)-based systems are representative work for solving information provision problems by implementing a specific platform [Masli et al., 2013, Hara et al., 2013]. Figure 2.5 (c) shows the positions of these types of HCI and CI studies. Since these studies do not further examine the impact of the new technology on the local information landscape of a community, there are opportunities to create a new problem space regarding the deployment of ICTs in communities.

2.3.2 Correlation and Causation between Community Features, LIL, and Community Outcomes

While the other categories of local community research have studied causations between community-related features, they usually focus on a small group of demographics or particular technologies. Unlike this kind of research, many studies from computational social science, political science, and sociology explicitly target community-level characteristics in understanding the dynamics of local communities. Their focus has been mainly to predict community features such as socio-economic status using computational techniques (e.g., machine learning) to provide low-cost, real-time indicators for various urban characteristics (Figure 2.6 (a)), and to understand the effects of community characteristics such as ethnic heterogeneity on other community outcomes (e.g., civic engagement) (Figure 2.6 (b)).

Community features predicted based on large-scale datasets include, but are not limited to, the socio-economic level in a city based on cellphone detail records (CDR) [Soto et al., 2011, Blumenstock et al., 2015, Lee et al., 2017], recognizability and socio-economic level using user-generated data from a web-based game [Quercia et al., 2013], land uses and points of interests (POIs) in a city using CDR

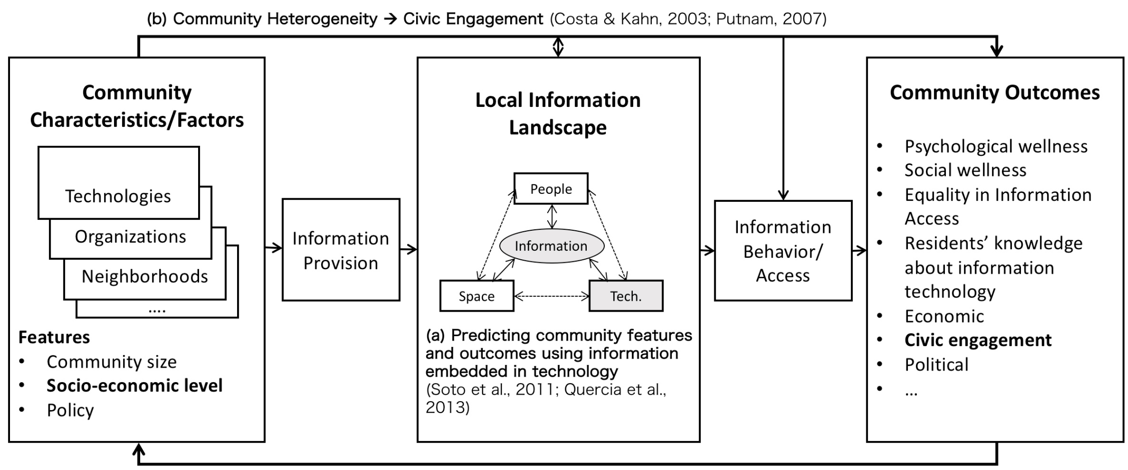

Figure 2.6: Mapping computational social science, political science, and sociology studies about local communities on the LIL framework.

and geo-tagged tweets [Frias-Martinez et al., 2012], human mobility patterns using Foursquare and GPS trajectories [Noulas et al., 2012], and local friendships depending on physical distance using geo-tagged Facebook data [Backstrom et al., 2010]. Often, these studies needed to conceptualize digitized local information for describing the fundamental characteristics of their raw data as a baseline when they computationally find patterns. In other words, this kind of prediction study usually assumed that local information was geographically and demographically biased due to the irregular distribution of geo-tagged data in a city and the technology users' inconsistent demographics [Quercia et al., 2013, Venerandi et al., 2015]. Although some studies suggest that demographic biases are not wide-spread in particular domains [Quattrone et al., 2015], it is a common understanding and challenge for urban computing researchers that geographical biases in geo-tagged data need to be adjusted using mathematical techniques and richer datasets [Wang et al., 2016a].

These studies, if viewed from the LIL perspective, develop prediction models that identify correlations between local information that exists in technical infrastructure (e.g., social media and cellphone vendors' databases) and community features such as socio-economic status. This research stream is not only methodologically useful due to the cost-effectiveness and mathematical novelty, but also theoretically meaningful in the LIL aspect because of the implications of local information itself. Since urban data is a type of local information available in a city, the efforts taken to use different types of data to predict community features can help understand the sources, distributions, qualities, and characteristics of various kinds of local information. Identifying biases in the data can reflect researchers' value on data quality (e.g., eliminating modeling errors) [Wang et al., 2016a]; but, it also could be the case that the geographical biases themselves may represent the characteristics of existing local information and have theoretical implications in their own right.

Meanwhile, studies that focus on causalities between community features and outcomes have contributed to social science theories [Head, 2007]. For example, civic engagement, a major community outcome, is a function of public interventions for connecting government and citizens [Brown and Keast, 2003] and of broader community features such as community heterogeneity [Costa and Kahn, 2003, Putnam, 2007, Laurence, 2009]. While similar to the studies that implemented ICTs in local communities for increasing civic engagement, these studies examined direct causal effects of community-level characteristics on civic engagement, which explains community-level dynamics rather than illustrating correlations between urban features.

2.3.3 Examination of Information Behavior in a Local Community

In library and information science, local information has often been studied from an information behavior perspective in research focusing on library uses. Since public and school libraries are known as important local information sources for diverse populations, studies about library use can be seen, to some degree, as being related to local communities. For example, important factors that shape individuals' library use and information behavior include cultural backgrounds [Liu and Redfern, 1997, Baron and Strout-Dapaz, 2001, Bordonaro, 2006], public libraries' efforts in technology education and assistance (i.e., digital literacy program) [Bertot et al., 2016], and the use of online community networks [Pettigrew et al., 2002]. Library-focused studies provide insight into the importance of institutional efforts to increase information access (e.g., IT training), but also have limitations in the geographical coverage of information sources by constraining their locations.

Other studies focused on immigrants' information seeking and use that were closely related to their adjustment and settlement in a city. Particular contexts and information sources were found as critical factors that affected immigrants' information behavior. These factors include close relationships [Quirke, 2012], the internet and online/offline social networks [Komito and Bates, 2011, Khoir et al., 2014], and word-of-mouth [Shoham and Kaufman, 2007]. Studies in everyday life information seeking (ELIS) further highlighted the exploration of what information practices (e.g., social media use) became part of particular populations' daily lives [Lingel, 2011, Sin and Kim, 2013]. These studies about information behavior regard-

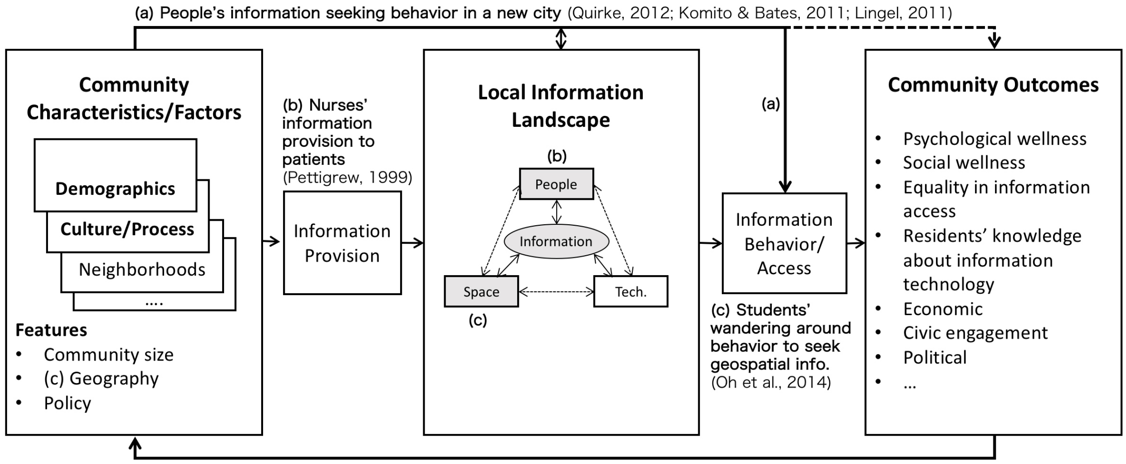

Figure 2.7: Mapping library and information science studies targeting local information on the LIL framework.

ing library use and daily life, if viewed through the LIL lens, deal mostly with the relationships between community features (e.g., library's programs, infrastructures, and cultural traits) and people's information behavior (Figure 2.7 (a)). Since information provision is a particular form of information behaviors, studies that focused on information provision can also be understood as a similar type of research to those mapped onto Figure 2.7 (a). From the LIL perspective, however, information provision research can be mapped onto the LIL framework differently since it distinguishes the materiality of information from the use of it (e.g., nurses' information sharing with seniors in a community clinic), thus is mapped to Figure 2.7 (b).

Similar to the studies on the ELIS of immigrants, Oh et al. [2014] studied international students' information behavior when they first came to the United States. Oh and colleagues found that newcomers' information seeking behavior, particularly their geospatial information seeking behavior, was unique in that people not only made use of technology and institutional information, but also depended on their social capital, ethnic groups, and the "wandering-around without any purpose" strategy to find information. The position of this study in the LIL framework is similar to the aforementioned ones (i.e., Figure 2.7 (a): the relationship between community features and information behavior). However, the wandering-around behavior in a city for acquiring information is noticeable, because this behavior exemplifies people's access and use of information directly through interactions with physical spaces. Along with the example of people's interactions with data-in-place such as posting on PosterVote [Taylor et al., 2015] that was intentionally installed for facilitating people's information sharing, the wandering-around behavior suggests that local information embedded on physical spaces actually shapes people's information behavior in local communities (Figure 2.7 (c)). Also, these studies show that the information people seek and use is not limited to that from information providers but exist pervasively in a community with *and* without respect to actors' intentions as the LIL framework indicates.

2.4 A Theory of Local Information Landscapes

As reviewed with research from different disciplines, the LIL framework (Figure 2.4) can explain various relationships between factors and characteristics of local communities. Particularly, it was possible to understand information provision as a process in the formation of local information landscapes by distinguishing it from other information behaviors with the assistance of the LIL lens. This initial assessment partially resolves the causality challenge of the preliminary model (Figure 2.1).

The mapping of empirical studies together suggests that the LIL framework is a meta-theoretical framework that explains diverse relationships between community characteristics, community outcomes, and local information landscapes. Therefore, it is possible to theorize these relationships found from the LIL framework as *a theory of local information landscapes (LIL theory)*[Lee and Butler, 2019].

While many empirical studies implicitly demonstrate some parts of the LIL framework, no research explicitly studied local information landscapes as a variable that affects or is affected by other community characteristics and outcomes. Thus, the causality challenge remains, and new problem spaces can be created to better understand the causal relationships between the components of the LIL framework. At the same time, this new model provides theoretical implications for understanding information inequality and information poverty from a material perspective.

2.5 Theoretical Implication: Information Deserts

Similar to the studies that take a socio-technical approach to solving information access issues, the notions of *information inequality* and *information poverty*, while having different nuances, conceptualize people's disparities in accessing information as socially-constructed phenomena [Schiller, 2013, Norris, 2001]. Information inequality has often highlighted social inequality imposed by such things as the power- and economy-driven provision of information in mass media [Schiller, 2013], uneven possessions of information by individuals in different social classes [Van Dijk, 2000], and the general ICT capacity of households [Hilbert, 2014]. Focusing more on the deprived side of the digital divide, information poverty often considers individuals' situational aspects such as discrepancies in peoples' digital literacy [Eshet-Alkalai, 2004] and their broader information processing incapability (i.e., information illiteracy) beyond the digital skills [Eisenberg, 2008]. Also, information poverty was understood as a result of the economic and technological deficiency [Haider and Bawden, 2007]. While the nuances and focus of this phenomenon are different across research communities, as Yu [2006] noted, it is an implicit assumption that most research studies on the topic of information inequality and poverty perceive information as a socially-constructed entity. By taking the LIL lens, it becomes possible to understand information inequality and information poverty from a material perspective.

Some notions that discuss material characteristics of information can be found in previous studies. Particularly, if digitized data on the internet is conceptualized, material aspects of the data could surface. For example, the concept of *data deserts* has been explored as a societal or disciplinary characteristic that is affected by the recent explosion in digitized data. In computational and human geography, researchers suggested that data deserts exist in technical systems where data is scarce because data availability and continuity differed before and after the big-data era [Kitchin, 2014]. In a similar vein, HCI researchers in the domain of volunteered geographic information (VGI) have studied the data creation process by volunteers and articulated the notion of *localness* that explains the origins of data and data providers [Sen et al., 2015]. This stream of research found that there were discrepancies between the locations of editors who contributed to geographically-specific

Wikipedia pages and the regions that these pages referenced, providing evidence of geographical inequalities in data creation. Rural areas were not only more likely to have data created by outsiders, but were also found to have lower quality data than urban areas on Wikipedia and OpenStreetMap due to the high portion of data created by automated agents [Johnson et al., 2016].

While these studies have limitations in understanding information inequality in general in a community due to their focus on digital data (leaving other forms of information behind), the term ,"deserts" in data deserts illustrates the meaning of materiality well. The use of the term, "deserts," has been prevalent in the discussion of *food deserts* that signifies people's disparities in accessing food resources such as grocery stores and restaurants [Walker et al., 2010]. Since food and the locations of food providers are material entities that have physical distances from each household, the concept of food deserts has been studied as a necessary condition of its manifestations in people's lives and food access. A takeaway from the food deserts studies is that researchers focused on the materiality of food as a factor that affects people's lives. Conversely, the social construction of food was also studied: a study that took a socio-technical approach in understanding food deserts examined the impact of families' socio-economic status on food's symbolic values [Fielding-Singh, 2017].

These two different understandings of food access issues are consistent with the two views on information access issues suggested in this study: socio-technical and material views. This observation and analogy justify the use of the term, "deserts" for conceptualizing the material aspects of information inequality or information poverty as *information deserts*. Based on the LIL theory, information deserts can be defined as structural and material states of local information landscapes that are pre- or necessary conditions of community-level information inequality (but not a sufficient condition). As López et al. [2014] suggested, there can be some material/structural forms of local information landscapes that could potentially affect people's information inequality. Some examples of information deserts are as follows:

- When local information is fragmented across different information sources [López et al., 2014].
- When local information is transient due to the material characteristics of embedding entities (e.g., word-of-mouth in a small group of people).
- When there is a lack of components of local information landscapes (e.g., lack of local bulletin boards in a neighborhood).

2.5.1 The Fragmentation of Local Information

When a certain type of local information is provided only to a part of the components of the LIL model, the distribution of information would be inconsistent across different information sources, and this might negatively affect people's information access. López et al. [2014]'s work is a representative example that demonstrates information fragmentation with respect to technical infrastructure owners. This structural characteristic is distinguishable from the concept of *fractured information landscapes*, suggested by Lloyd [2017], which takes a socio-cultural view on the fracture of information landscapes where an individual's familiarity with information sources and contexts is a key in understanding the overall landscape.

The uneven distribution of local event information embedded in different technical infrastructures (e.g., neighborhood blogs and city-wide newspaper websites) shows that certain types of event information (e.g., museum-initiated events) are not available from neighborhood-based websites; instead, they are available in a city-wide information source. This implies that information deserts can be unintentionally created as a byproduct of organizational strategies or due to the nature of particular information types. From a user perspective, it is possible that an individual would not be able to access every information available in the community unless he or she looks into all the information sources. Thus, the high fragmentation of local information could be potentially detrimental to people's information access at the community level (c.f., it does not represent information inequality by itself; rather, it is a necessary condition of information inequality).

2.5.2 Transience of Local Information

Even if information were provided evenly to the information sources in a community (so there is no fragmentation), the transient characteristic of the entity that embeds information could create another form of information deserts. For example, if a piece of information was provided to a couple of people through word-of-mouth and they forgot about that information, it is understood from the LIL view that information deserts are created due to the transient characteristic of information (because this information is no longer accessible). Another example can be found in a case where information is in a volatile memory of a computing system. If, for some reason, pieces of local information are stored only in a Random-Access Memory (RAM) that stores data while power is on and there was a sudden power outage for a few seconds, then that information disappears due to the characteristics of RAM.

The transience of information that stems from the material characteristics of the embedding components has a different meaning from that of the information that is available at the moment of practices, i.e., contingent information, such as medical information acquired at the moment when a nurse examines a patient's body [Bonner and Lloyd, 2011] or that is informal/contextualized, i.e., vernacular information [Trace, 2008]. This type of information emerges from collective group activities, cultural experiences, and an internal view on information, thus, is called local knowledge as well [Lloyd, 2014]. Contingent or vernacular information is available and sought depending on people's understanding of situated practices and is perceived differently by time, space, and individual. Unlike this type of information, information deserts created based on the transience of information conceptualize any material states with a lack of information that stems from the characteristics of embedding entities.

2.5.3 Lack of Components of the Local Information Landscapes

If there are not enough infrastructures, people, or spaces that can contain and provide local information, information deserts can be created as well. For example, physical bulletin boards are spatial entities that embed local information. People would have a hard time finding certain types of local information from locations where bulletin boards are scarce or absent (e.g., neighborhoods with few numbers of coffee shops and public libraries) by eliminating the capability of information creation/provision in the local community. Similarly, a lack of neighborhood gatherings or technical infrastructures is another form of information deserts created by limiting the possibility of information provision.

2.6 Research Agenda

Given the forms of information deserts and LIL framework with respect to the scale, complexity, and persistence of the LIL components, research studies that can be designed in several academic fields are presented by identifying gaps using the LIL framework (Figure 2.4).

2.6.1 Human-Computer Interaction (HCI) and Community Informatics (CI)

A few local community studies from HCI and CI highlighted information provision by focusing on the amount of information disseminated through infrastructures and people [Carroll et al., 2011, Foth et al., 2011]. However, the effect of the information provision is still unexplored thoroughly for the three types of provisions (i.e., provisions to technical infrastructures, people, and physical spaces). Additionally, HCI and CI tend to focus on one or a few particular systems when considering information provision. Mapping empirical studies onto the LIL framework (Figure 2.5) suggests that the impact of implementing new technology can be further explored by considering new categories of dependent variables: the characteristics of local information landscape and people's information behavior regarding them.

In most cases, some information systems, either technical or social, already exist when an ICT is implemented. Even for a developing country, for example, social systems such as a word-of-mouth network or local communication structures exist for exchanging information within a community. Designing and implementing a new ICT may help residents communicate better with each other regardless of the existing structure of local information landscapes, but it is also possible that adding one more system may lead to the high complexity of residents' information access practices. Conducting user studies using only a new system may not capture the effects of adding new technology. Accordingly, new categories of research studies can be designed by considering these effects:

- Comparing people's information behavior before and after deploying a new system given conventional information sources (by focusing on their daily experiences, rather than their behavior regarding a new system).

- Identifying the amount and types of information that can be complemented in a city by implementing a new system.

- Designing and developing visual analytics tools that show how information deserts in a city change over time before and after implementing a system (e.g., how does the degree of fragmentation in LIL change?).

- Identifying and justifying the position and role of a new system within the existing information landscape of a city - whether it is an implementation of a complementary system that provides an easy access structure [Hu et al., 2013] or a completely new system that provides new information that was hardly provided before.

2.6.2 Computational Social Science and Political Science

Many researchers have focused on identifying and adjusting geographical biases from datasets [Wang et al., 2016a], because it was essential to verify the characteristics of geospatial data and their findings in justifying prediction models. If we think of the geospatial data as a form of local information that is embedded in technical infrastructure, it is possible to say that biases in the data may actually represent the structure of local information landscapes (or potentially information deserts) by themselves. In this sense, quantifying and explaining diverse aspects of information deserts using geo-tagged data would be possible for future studies. Also, studies that examined relationships between community features and outcomes (e.g., the effects of community heterogeneity and civic engagement) can be re-visited by taking the LIL framework into account. Examples are as follows:

- Constructing mathematical models that quantify:
  - The degree of fragmentation of local information in a city.
  - The availability of each component of the LIL model.
- Predicting the effect of adding an information source in a city (e.g., how can we measure the impact of adding a Twitter-based local information application on a city's information deserts?)
- Studying causal relationships between information deserts (e.g., the complexity of local information infrastructures) and community outcomes (e.g., socio-economic status).
- Studying how the LIL features (e.g., the fragmentation of local information) mediate or moderate the relationships between community features and outcomes.

2.6.3 Library and Information Science

Existing theories about information access and behavior can be extended by making use of the LIL theory. Also, based on the library and information science studies mapped in Figure 2.7, it is possible to identify some hidden patterns of information behavior regarding the complexity, scale, and persistence of the LIL components. For example, since some studies about geospatial information embedded in physical spaces and social systems focus on particular contexts and scales of LIL components (e.g., space as a city environment) [Oh et al., 2014, Pettigrew, 1999], the LIL theory can help find other features of the LIL components by diversifying the complexity, scale, and persistence of each component. Some examples of such opportunities are as follows:

- Exploring people's information seeking behavior where the scale of spatial entities is small and transient (e.g., customers' information seeking behavior in a coffee shop where the turnover rate of flyers on the bulletin board is fast).
- Studying how the interplay between multiple sources that contain the same information manifests in different contexts (e.g., when the same information is available both from a bulletin board and librarians in a public library, how each affects people's information seeking behavior; how each of these components extends or limits the capability of the other; and how the information redundancy plays a role for library visitors in accessing it).
- Exploring how and why information is provided to a physical space (e.g., how information about local events, such as farmers' markets, becomes embedded on physical spaces, and why? What are the implications of actors' information provision behavior?).

2.7 Discussion and Limitations

Despite the advantages of the proposed theory and new research opportunities derived from it, some issues remain. First of all, the complexity of each component is still largely encapsulated in an abstract form. The component of *people*, even from a material perspective, cannot be completely separated from human factors since some characteristics such as memorizing ability and their attentions are important in keeping information. The interactions and dynamics of groups and organizations themselves are very complicated processes that can affect the form of local information substantially. Although existing theories and models in HCI and Psychology explain diverse aspects of information users well, the *people* component needs to be further explored from the material perspective to characterize the local information landscape better. Articulating the scopes and priorities of the characteristics of the LIL components would make this model more useful for local community scholars.

Additionally, more empirical studies from diverse academic disciplines need to be explored to better understand and refine the LIL framework. In this chapter, only some of the major academic fields that study local communities and cities were reviewed to construct the LIL theory. However, there are many other empirical studies and academic fields that can complement and assess the LIL theory (e.g., urban and transportation studies are not reviewed in this chapter). More efforts are needed by scholars from diverse academic fields to fill the potential gaps between studies that examine community-level characteristics.

Lastly, the LIL model has limitations with regard to explaining the emergence of highly-situated information in the local community context. As Lloyd [2010] noted, information and knowledge often arise from people's collective understanding of physical environments and practices. By focusing on the materiality of information, however, the LIL model is more useful for explaining the features of standardized/legitimized public information than those of nuanced, situation-dependent information. This focus could be seen as a trade-off between the scalability and scope of the ontological model. Further research and case studies are needed to understand these tensions and trade-offs better and to refine the LIL framework.

Despite these limitations, this work is meaningful in several aspects. While assessing the LIL framework, it was possible to review a wide spectrum of local community research across disciplines beyond the scope of information science. It is our hope that scholars from different fields communicate with each other more by using and developing the LIL theory with studies that fill in the empirical gaps.

Distinguishing information provision and behavior/access is another contribution of this chapter. While information creation is not a new concept in the field of information science [Trace, 2007], information provision to three material components of the LIL model has rarely been theorized in the context of local communities. Through the LIL theory, scholars who study information creation can benefit from systematic approaches to thinking about the ways that information is provided. Also, information behavior studies can benefit from the ecological forces of local information landscapes that shape individuals' information behavior and access (beyond the discussion of how created information exists).

Chapter 3: Study Design: Information Deserts and Community Characteristics

3.1 Rationale

As reviewed in the previous chapter, the LIL framework explains diverse relationships between different community-level features and individuals' information practices. However, no study has explicitly examined the impact of information deserts or local information landscapes as a community-level factor. To provide an assessment of LIL theory, it is necessary to empirically study the effects of information deserts on other community characteristics. Among many other community-level qualifications in local communities, *community engagement* is one of the most important characteristics on which researchers in community informatics (CI), human-computer interaction (HCI), communication, sociology, and political scientists have focused [Han et al., 2014, Putnam, 2007, Costa and Kahn, 2003].

Community engagement, while there is not a universal definition, is the process and practice of collaboration among people who share similar goals and values, particularly among those who are associated by geographical proximity [Tindana et al., 2007]. Individuals' engagement in social systems provide opportunities to acquire information and work together to solve problems [de Montjoye et al., 2014]. For example, community participation and awareness can positively affect shared problem solving and physical/psychological well-being [Zakus and Lysack, 1998, Helliwell and Putnam, 2004]. This is not only true in team work and business settings, but also in everyday-life contexts where technological systems are often designed as part of infrastructuring processes in communities [Grudin and Pruitt, 2002, Korn and Voida, 2015], which in turn makes a local community resilient.

There are many different factors that affect community engagement. Social capital, one of the indicators for community engagement and a necessary condition for their civic participation, is a basis for healthy society [Gil de Zúñiga et al., 2012]; also, social capital is known as a strong factor that induces civic participation by reducing corruption rate in political contexts [Nannicini et al., 2013]. This is why communication scholars have also focused on information and communication infrastructures as a key means to provide better access to storytelling community resources [Kim and Ball-Rokeach, 2006]. Additionally, the level of community engagement can be affected or mediated by technological or human factors such as the deployment of information systems, technology use patterns, and digital literary programs [Abd Aziz et al., 2015, Han et al., 2014, Head, 2007]. Civic engagement, a particular form of community engagement, is a function of public interventions for connecting government and citizens such as network arrangements [Brown and Keast, 2003], and of broader community features such as community heterogeneity [Costa and Kahn, 2003, Putnam, 2007].

While the relationships between these factors and community engagement can be mapped onto the LIL framework (Figure 2.4), it is still unclear what role local information landscapes play with respect to these known relationships. With the assistance of LIL theory, it is reasonable to imagine that information in a local community is available in technological infrastructures, people's interactions with local organizations/government agencies, and their mobility patterns with regard to physical information sources. Also, it is possible to hypothesize that the information provided to diverse entities affects people's participation in community gatherings and discussions, because people need to somehow acquire the information about the meet-ups before they attend any events. Especially in recent decades, there have been radical changes in local information landscapes, which include, but are not limited to, the emergence of the internet websites as popular information sources and a decrease in local newspapers [Gao et al., 2018]. This suggests that local information landscapes might have played an important role for people's gatherings and activities.

As an initial exploration of this problem space, this chapter presents a study that examines the effects of the fragmentation of local information, a dimension of information deserts, with respect to other community characteristics. Specifically, this study focuses on information about local events, such as farmer's markets and local concerts, as the context to empirically study the role of local information landscapes in the process of people's community engagement. Community members' strong relationships with others, or social capital, can be established based on interpersonal trust through online and offline activities, which is one of the driving forces that shapes people's engagement [Brehm and Rahn, 1997]. Because maintaining both online and offline relationships gives rise to tie strength between community members, a necessary condition of community engagement [Sessions, 2010], the local events landscape as the study context is meaningful and plausible for examining the relationship between local information landscapes, community engagement, and other community characteristics. Through identifying the roles of information-driven factors, it is possible to understand the process of how information deserts are created in local communities and what their impacts on these communities are.

To begin the study, findings about the relationship between community heterogeneity and community engagement are reviewed to provide a baseline theoretical framework upon which this study builds. This chapter then discusses why local information landscapes would play an important role within the existing theoretical relationships. Also, based on the literature review about the concepts of cultural diversity and related work, this study defines *cultural activity diversity*, the extent to which a manifestation of a community's cultural activities is diverse, which is distinguishable from the concept of cultural diversity that has been studied in previous work. Through this stream of the literature review and inferences, research questions and hypotheses are suggested in the last section of the chapter.

3.2 Community Heterogeneity and Community Engagement

Researchers in community informatics have studied the impact of *information and communication technology (ICT)* on people's engagement as an important factor [Han et al., 2014, Kavanaugh et al., 2014]. However, these studies usually focus on individual- or group-level attitude changes rather than on community-level characteristics. Although community-level engagement or social capital cannot be explained without considering individual- or group-level dynamics as Costa and Kahn [2003] explains, measurements and approaches to studying these two distinguishable, but related, phenomena are usually different in terms of the scope of study subjects, specificity of study contexts, and instruments/protocols. Because the focus of this dissertation is on understanding community-level dynamics so that key variables (i.e., information deserts indicators) can be understood and monitored by change agents such as policy makers, empirical studies that focus on community-level/structural features in a broad context, rather than individual-level/functional features in a particular context, would provide better implications for theoretical constructs regarding LIL theory. As such, community engagement in this study is treated as a characteristic of a local community or city, rather than an individual or group characteristic.

As mentioned above, there are many factors that affect people's community engagement. Among them, community heterogeneity has been studied extensively by sociologists, political scientists, and economists due to its major effects on community performances in varying contexts [Putnam, 2007, Costa and Kahn, 2003, Abascal and Baldassarri, 2015a]. Especially because of the high ethnic/cultural diversity and increasing number of immigrants in the United States, the social impact of community heterogeneity has been one of the most important topics on which social scientists have focused for decades [Dinesen and Sønderskov, 2018].

While a motivation of researchers in studying community heterogeneity is often on the high international mobility and inter-racial/ethnic dynamics within and across U.S. cities, community heterogeneity itself is not a single construct but a multi-dimensional concept that includes, but is not limited to, ethnic heterogeneity [Costa and Kahn, 2003], cultural diversity, [Bail, 2014], and heterogeneity in a community's economical/financial structures [Benabou, 1996]. For example, Putnam [2007] argued that, in the short run, ethnic heterogeneity negatively affects social solidarity and trust between local residents. Conversely, this study also indicated that ethnic diversity can be positive in the long run for community engagement. This finding, particularly the findings for the short run, has been supported by many different studies in sociology, political science, and economics. Costa and Kahn [2003] confirmed, using datasets from the United States, international, and historical contexts, that homogeneous communities in terms of ethnicity and socio-economic status foster greater levels of social-capital production. These findings are relatively consistent in different temporal and spatial contexts, with varying dimensions of community heterogeneity between cultural, racial, ethnic, and income diversity/inequality [Alesina et al., 2004, Alesina and Ferrara, 2000]. When race and nationality are homogeneous within a community, trust tends to be high [Glaeser et al., 2000]. Cultural similarity in terms of their lifestyle and identity was also found to reduce loan default rates in Peru [Karlan, 2005]. These findings suggest that different dimensions of community heterogeneity might not be orthogonal, but somehow related to each other.

Although many studies find that community heterogeneity negatively affects community engagement, trust, and socio-economic status, there are also contradictory findings, even with similar datasets or contexts [Schaeffer, 2013, Meer and Tolsma, 2014, Gereke et al., 2018]. Abascal and Baldassarri [2015a] reexamined Putnam [2007]'s work using the same survey datasets that Putnam used after reconceptualizing heterogeneity and combining some census data. They showed that ethnic heterogeneity has minimal or negligible effects on civic engagement when considering different perceptions of ethnoracial groups toward others. According to their study, it is actually the particular ethnic group size/dominance that shapes social trust, not the ethnic heterogeneity *per se*. This study concludes that some of the previous work on the effects of community heterogeneity mis-conceptualized "social capital" without integrating the *intergroup contact/conflict* literature [Pettigrew, 1998] and overlooked the inconsistent nature of inter-ethnic dynamics (e.g., "community heterogeneity" has different meanings for white communities and non-white communities). Using a survey dataset from Canada, Stolle et al. [2008] provides similar implications that social ties, that is, actual inter-ethnic contacts and their strengths, mediates the relationship between community heterogeneity and social trust within the community.

Many studies illustrate these inconsistent findings on the advantages and disadvantages of community heterogeneity as a conflict between contact theory [Pettigrew, 1998] and conflict theory [Tajfel et al., 1979]. While contact theory explains the favorable effect of community diversity because inter-group contact reduces people's prejudice toward out-groups, conflict theory explains its negative effect on trust based on actors' perceived threat toward out-groups. These inconsistent findings on and different conceptualizations of community heterogeneity suggest that there might be certain intermediary processes between community heterogeneity and community engagement or other contextual differences between study domains. Particularly, ethnic or racial diversity, one of the community characteristics, might have other important processes associated with them in shaping the discrepancies in social ties, trust, and people's baseline perceptions toward in-groups and out-groups. In other words, if using the terminologies in LIL theory, oversimplifying the manifestations of a community's material structures/forms in studying community dynamics could lead to biased interpretations of social phenomena, even if analytical models have been specified correctly [Abascal and Baldassarri, 2015a].

Previous work has overcome the limitations of using simplified and aggregate community characteristics in understanding local communities by (1) focusing on individual-level characteristics by including their demographic/social characteristics into account, (2) distinguishing baseline characteristics of different groups toward others, (3) including important control variables, such as socio-economic inequality and the dominance of particular social groups, in the analytical model, and (4) including confounded factors in the analytical model to distinguish the main effect from them [Gereke et al., 2018, Abascal and Baldassarri, 2015a, Sturgis et al., 2014]. Accordingly, many studies re-examined the negative relationship between ethnoracial heterogeneity and social trust/tie by including other socio-demographic and socio-economic factors such as socioeconomic deprivation, age, and gender [Sturgis et al., 2014, Stolle and Harell, 2013, Laurence, 2009]. For example, Sturgis et al. [2014] included age, gender, and economic deprivation in the individual-level analysis and found that socio-economic deprivation is confounded with ethnic diversity in London, U.K., because immigrants and people of color tend to be in a lower economic status compared to long-standing residents. In the U.S. context, this confounding effect between ethnoraical heterogeneity and socio-economic deprivation was studied in various ways, not only based on the national survey data collected for decades, but also with redlining maps data which showed that the racial zoning of the 1920's and 30's still predicts the poverty level in modern neighborhoods [Mitchell and Franco, 2018].

Age was also found to be a moderating factor between community heterogeneity and social trust because social contact is predominant among young groups while social conflict due to lack of interaction is higher among old groups [Laurence, 2009, Sturgis et al., 2014, Stolle and Harell, 2013]. Laurence et al. [2019] further dove into this relationship with respect to conflict theory and concluded that perceived threats against out-groups moderate the effect of ethnoracial diversity on social cohesion, rather than mediating the relationship (so, people's attitude can change over time). Therefore, this study first examines whether these known relationships still stand in the context of local events and meetups in U.S. cities (beyond people's civic participation and perceptions).

- RQ1: How is the ethnoracial heterogeneity related to socio-economic deprivation, socio-demographic characteristics, and people's participation in local activities?
  - RQ1-1: How does ethnoracial heterogeneity affect community engagement?

- RQ1-2: How does socio-economic deprivation affect community engagement?
- RQ1-3: How does ethnoracial heterogeneity affect socio-economic deprivation?
- RQ1-4: Do socio-demographic features (e.g., age) moderate the effect of socio-economic deprivation on community engagement?

Studying these research questions in the local event context is meaningful not only because of its contextual uniqueness, but also because of its practical usefulness. The majority of previous studies on this topic relied on survey questionnaires and census data. However, it is challenging to measure people's trust level and social tie strengths at the community level because of the inaccuracy of self-reporting and sampling issues (e.g., it is not easy to collect survey data in a randomized setting). While social media data also has some limitations in its provision and user demographics, using large-scale data that is collected from online communities provides an alternative way of capturing the manifestations of social ties and trust. Once this baseline relationship is examined, the next step is to study the effects of information deserts to extend the theories of community engagement.

3.3 The Fragmentation of Local Information

This study argues that, in addition to the strategies and granular conceptualizations to reduce oversimplification biases, identifying additional intermediary processes can further uncover community-level dynamics. Particularly, this study shows that known intermediary processes such as inter-group dynamics and socio-economic forces might be closely related to the local information landscapes, because (1) information sources that provide information about meetups and local gatherings would foster diverse populations' face-to-face interactions (thus, high inter-group contacts), (2) the contents of local information increase the visibility of and perceptions toward other cultural/ethnic groups in the community (thus, lowering the psychological barriers and perceptional threats toward out-groups), and (3) demographic/social groups' disparities in accessing the information sources would play a role in determining the composition of participants in local meetups (either due to structural factors or individual's technology literacy).

With respect to LIL theory, there can be several different kinds of community-level features that comprise information deserts, such as insufficient provision of information, the fragmentation of local information across different sources, and lack of entities that embed information [Lee and Butler, 2019]. Among these, as the first step to explore this problem space, this study focuses on the fragmentation of local information because this phenomenon has been rarely studied compared to other dimensions of information deserts (e.g., low broadband penetration rate is an instance of lack of entities that embed information [Czernich et al., 2011, Grubesic, 2004]). As partially demonstrated by López et al. [2014], the fragmentation of local information is one of the major causes for information deserts (i.e., the structural or material states that are necessary conditions of information inequality), because it decreases opportunities for information users to find information if they use only a few information sources. Based on their study targeting Pittsburgh's information sources, López et al. [2014] argued that information users will have to use more than a few of information sources to find desired information in the city because any of the information sources provide no more than 20% of the entire information available in the city; this percentage would be less so if local information were highly scattered across different sources.

In the field of communications, a similar concept has been introduced. The concept of *media fragmentation* and its operationalization have been widely studied, because identifying the audience distribution of mass and social media is a key to understanding the relationship between people's selective media exposure/consumption and their behavior/perceptions [Thorson and Wells, 2015, Taneja and Mamoria, 2012, Aalberg et al., 2013]. Since seeking and using information through diverse media can be perceived as "consumption," and often related to marketing dynamics, the distribution and choices of media users were of interest to marketing researchers as well [Wilbur, 2008]. From an information science perspective, these studies are theoretically meaningful in understanding information users' behavior on a large scale because the audience fragmentation could be one of the *social polarization* dimensions [Webster and Ksiazek, 2012]. In particular, the audience fragmentation approach in the contexts of news and advertisements has potential to benefit information behavior researchers by introducing a new community structure that is related to individuals' information behavior. This suggests that, similar to community engagement or social equity, audience fragmentation in the media landscape is a community feature that might affect or be affected by local information landscapes, rather than an alternative to the fragmentation of local information (Figure 2.4).

Meanwhile, the fragmentation of local information itself has been rarely studied as a community feature or factor in relation to other community characteristics. If socio-economic deprivation is high, people have less trust and social cohesion (thus, social segregation), which might lead to uneven provision of local information to different information sources (i.e., there might be few shared knowledge management systems between different socio-economic groups). Conversely, if poverty level is very high, it is also possible that low-income individuals might not be able to provide enough information to technological infrastructures consistently, or cannot even organize local events, due to the time and cost of using technology for their personal interests. This is an open question and is worth studying how socio-economic status affects the provision of information, which results in the fragmentation of local information.

- RQ2: How does socio-economic status affect the fragmentation of local information?

If the fragmentation of local information is high, as mentioned above, people will have less opportunities to learn about available information compared to a situation where each information source provides a sufficient amount of event information. This indicates that the fragmentation of local information might be closely related to community engagement as well.

- RQ3: How does the fragmentation of local information affect people's participation in local activities?

3.4 Cultural Activity Diversity

People's participation in local events can indicate the extent to which they engage in local activities, discussions, and neighbors. At the same time, their on-going interactions with in-groups/out-groups, urban spaces, technologies, and social norms would co-shape the community's local culture and social structures. Local actors, such as local businesses and residents, usually organize events with structures (e.g., payment process, group norms, and membership structure). In other words, social agents surrounding an event, such as event attendants and local vendors, *draw upon* these structures and enact within the social boundaries. Meanwhile, these agents' activities re-shape the event's social structures, according to the structuration theory [Giddens, 1986].

The collection of local activities is a result of people's cultural/ethnoracial backgrounds, the local community's customs, event organizers' philosophy in organizing events, organizational structures that are derived from their philosophy, and the complex interactions of these factors within the spatio-temporal space of the social system. The manifestation of such interactions is also an important dimension of local dynamics that might have something to do with local information landscapes, community engagement, and socio-economic status. Because the collection of local activities represents part of a local community's culture, the diversity of activities is called *cultural activity diversity* from a community perspective in this study. This community-level characteristic might be more fluid than ethnoracial heterogeneity and subject to change over time relatively quickly.

Cultural diversity in the literature is closely related to community heterogeneity, socio-economic deprivation, and demographic characteristics [Hristova et al., 2018, Uslaner, 2018, Park et al., 2015, 2017]. This is because, when referring to *cultural diversity*, many studies focus on people's national and ethnic backgrounds because of one's high dependency on his/her underlying traditions, behavioral patterns, and value systems that are embedded in nations and ethnic groups when it comes to social activities and utterances [Ottaviano and Peri, 2006, Babacan, 2005]. Representative work of country- and ethnic group-based approaches to understanding individual's cultural background is Hofsteds [1980]'s cultural index that quantifies each country's cultural dimensions based on how social norms and collective values manifest in people's perceptions and daily lives. This kind of measure has been extensively used in studies that examined cross-national in behavioral and social dynamics [Park et al., 2017, Mazanec et al., 2015, Pedrini et al., 2016].

Because the term, "cultural diversity," has been frequently used as an alternative or in relation to national and ethnic backgrounds, when scholars point to social/physical manifestations of culture such as people's activities and artifacts, they often used other terms such as *cultural tastes*, *cultural omnivorousness*, or *cultural specialization/capital* [Peterson and Kern, 1996, Hristova et al., 2018, Bail, 2014, Bail et al., 2019]. In other words, these terms are used for describing the cultural characteristics of individuals, communities, or their enactments in a conceptual space where the producers and audience of cultural artifacts co-exist. Research that focuses on this aspect of cultural characteristics has been often studied in particular contexts such as the music consumption market [Park et al., 2017], online search patterns, and place-driven characteristics in urban scenes [Hristova et al., 2018].

Some of these studies are based on individuals' cultural tastes and behaviors and are often operationalized based on the collective understanding of individuals' boundary spanning behaviors between different cultural genres (i.e., how many types of cultural products are consumed by individuals and how various their tastes are across different types), which scholars call *variety* from an *audience* perspective [Bail, 2014, Goldberg et al., 2016]. Conversely, as Goldberg et al. [2016] suggest, organizational theory scholars have often focused on *atypicality* from a *producer* perspective which conceptualizes the extent to which a cultural product/artifact is different from existing classifications and genre systems (e.g., [Dubois, 2012, Pontikes and Hannan, 2014, Park et al., 2017]).

Viewing from these two lenses, local events are a collection of different localized markets (from a producer perspective) and social groups' activities (from an audience perspective) that range from businesses to interest groups to political movements. While the amalgamation of local activities and meetups provides senses of both the variety of consumers' contextualized, personal tastes (i.e., cultural omnivorousness) *and* producers' strategies for organizational success in the ecology of cultural/social activities (i.e., cultural typicality/atypicality), it is not quite accurate to conceptualize the entirety of activities using one of these notions. Therefore, this study focuses on *cultural activity diversity*, rather than cultural diversity, that connotes the variety of local activities that abide by established genre systems from a participant perspective and, at the same time, the level of atypicality that is shaped by event organizers' strategies.

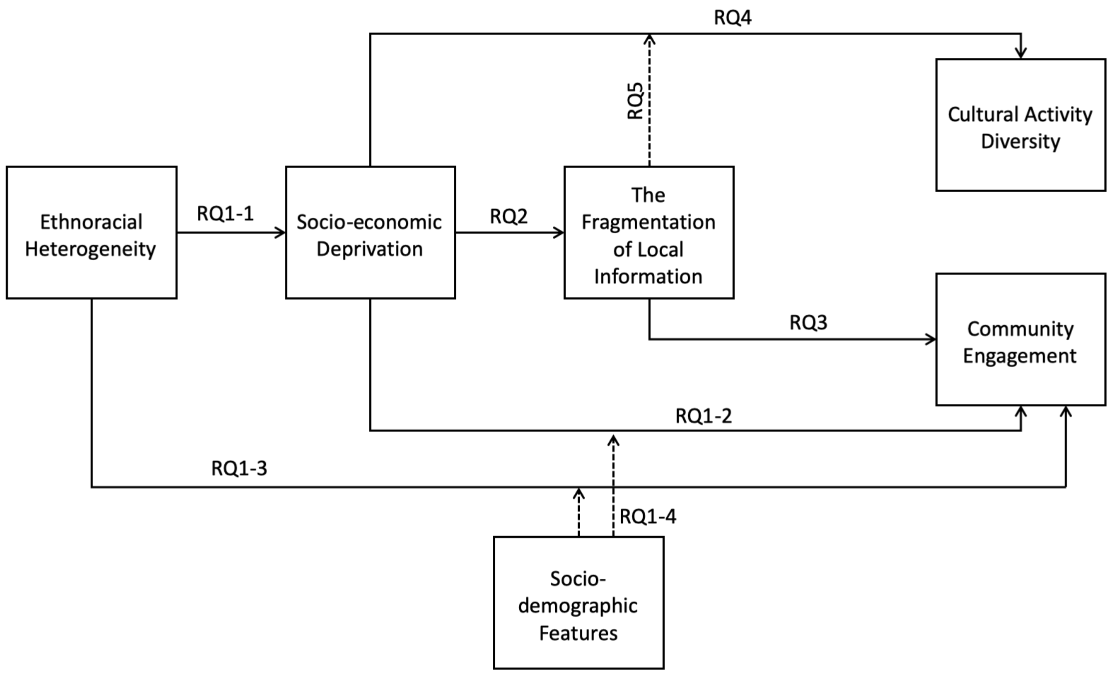

Figure 3.1: A theoretical relationships diagram for research questions.

The previous work provides evidence about the effects of people's cultural backgrounds on civic engagement [Putnam, 2007, Abascal and Baldassarri, 2015a], cultural specialization/capital on urban development [Hristova et al., 2018], and cultural backgrounds/values on cultural openness (indicating that technology globalization does not lead to cultural convergence across countries) [Park et al., 2017]. However, no research, to the extent of my knowledge, has studied cultural activity diversity as a variable on a large scale with respect to other community characteristics. Studying this new construct is meaningful with regard to community heterogeneity, engagement, and social cohesion because it (1) conceptualizes a holistic landscape of cultural activities that happen in a community rather than a particular scope of the cultural context, (2) focuses on people's face-to-face interactions rather than their online behaviors, which provide better implications for the theoretical framework from a social contact perspective, and (3) sheds light on the structuration process of the social system in which the complexity of local event structures and local agent heterogeneity co-exist, through the analysis of actual manifestations of local gatherings and activities.

Based on these rationales, it is reasonable to ask questions such as whether and how socio-economic segregations/deprivations are related to the diversity of local gatherings/meetups and how this relationship is affected by the fragmentation of local information. The research questions for this study are illustrated in Figure 3.1.

- RQ4: How does socio-economic deprivation affect cultural activity diversity?
- RQ5: How does the fragmentation of local information moderate the relationship between socio-economic deprivation and cultural activity diversity?

3.5 Hypotheses

RQ1 is intended for reproducing or replicating previous findings on the relationship between ethnoracial heterogeneity, socio-economic deprivation, socio-demographic features, and community engagement in the context of local events, as reviewed in Section 3.2, which will provide baseline statistics and ensure data consistency. Because people's participation in local events such as neighborhood gatherings and farmer's markets is an indicator of community engagement, which is directly associated with people's trust, cohesion, and social ties, it is possible to hypothesize that socio-economic deprivation that is known to be confounded with ethnoracial heterogeneity would negatively affect people's participation in local events. As such, hypotheses H1-1, H1-2, and H1-3, as a whole, examines whether ethnoracial heterogeneity is confounded with socio-economic deprivation.

- H1-1: Ethnoracial heterogeneity would be positively correlated with socio-economic inequality/deprivation (RQ1-1).

- H1-2: Socio-economic inequality/deprivation would be negatively correlated with people's participation in local events (RQ1-2).

- H1-3: Ethnoracial heterogeneity would be negatively correlated with people's participation in local events (RQ1-3).

- H1-4: Socio-demographic features would moderate the effects of community heterogeneity on community engagement (RQ1-4).

- H1-5: Socio-economic inequality/deprivation would mediate the effects of community heterogeneity on community engagement (RQ1).

If socio-economic deprivation is high, it will be more likely that the community is segregated and different socio-economic groups will make fewer contacts with each other. If this is the case, different groups of people would tend to provide local information in ad-hoc manners without communicating with each other, which in turn creates a highly fragmented information landscape in the city.

- H2: Socio-economic deprivation would be positively correlated with the fragmentation of local information (i.e., making it more fragmented) (RQ2).

If the fragmentation of local information is high, it is more likely that people have less exposure to available information in their community in a given amount of time because each information source will provide only part of the local information, which in turn, will negatively affect their participation in local events. This leads to the hypothesis that high fragmentation of local information would negatively affect community engagement.

- H3: The fragmentation of local information would be negatively correlated with community engagement (RQ3).

When socio-economic deprivation is high in a community, residents will be more likely to have fewer resources to organize local events. As a result, it might be the case that organizations and a few individuals who have resources, rather than most of individuals, are the majority of the event organizers in a highly-deprived community, making the variety of local events low. Therefore, it is possible to hypothesize that higher deprivation would make the community less diverse in terms of their activities. If people can be exposed to information about diverse activities, conversely, people would be able to attend existing events more, and the negative effect of socio-economic deprivation might be lessened, moderated by the fragmentation of local information.

- H4: Socio-economic deprivation would be negatively correlated with cultural activity diversity (RQ4).

- H5: The fragmentation of local information would negatively moderate the effect of socio-economic deprivation on cultural activity diversity (RQ5).

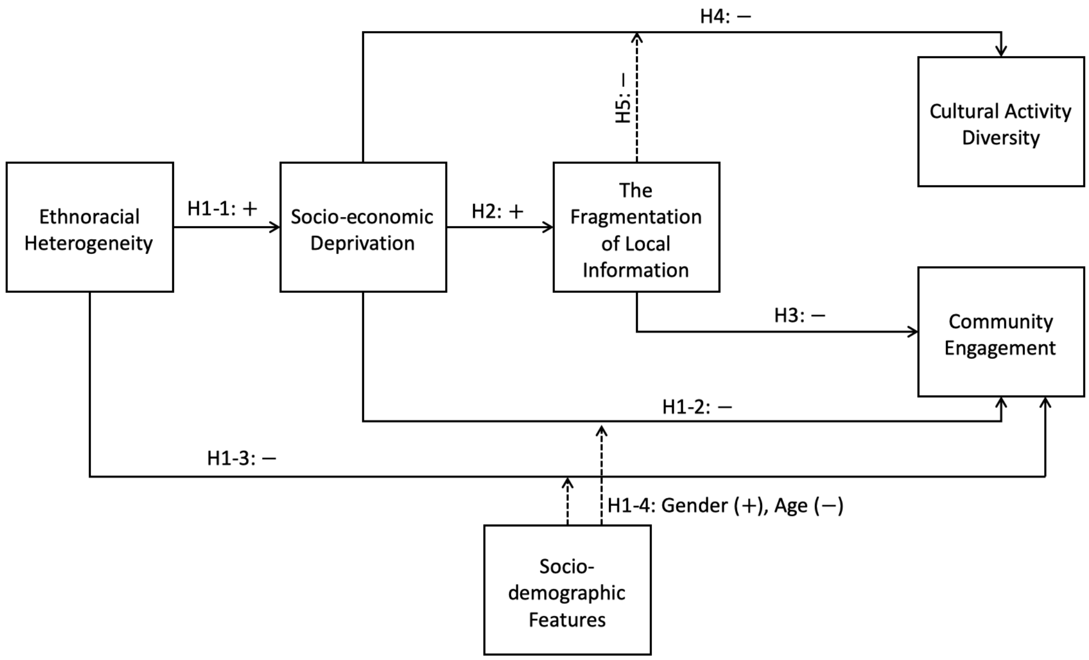

Figure 3.2: Hypotheses for the empirical study.

Overall hypotheses are depicted in Figure 3.2. These hypotheses are tested in a quantitative study using data from a sample of major U.S. cities. This study uses local event data that is collected from several information sources, American Community Survey (ACS) from U.S. census, and geospatial data for segmenting regions.

Chapter 4: Computational Approach to Measuring the Fragmentation of Local Information

Local information landscapes, especially those that are formed based on technological infrastructures, are more likely to affect people's daily lives in a metropolitan scale, rather than in a municipal neighborhood or country scale. The reason is because people's daily mobility and activities are often limited to a metropolitan scale rather than a neighborhood or national scale. For example, average commute distances for one- or two-member households in U.S. cities varied from 24 km (15 miles) to 80 km (50 miles) in 2009 [Zhu, 2013]. This indicates that people's mobility and interests in their daily lives tend to range within a metropolitan area. Due to this reason, "urban area" is chosen as the geospatial unit of analysis.

If not using the national survey and census data, it is usually challenging to collect city-level data for different cities at scale due to inconsistent data sources, high cost in collecting data, and regulatory limitations. Another challenge is that target variables, especially multi-dimensional constructs, are often not available from such data sources. Because of these practical and theoretical reasons, this study uses online data, computational modeling, statistical inference, and machine learning

(ML) to cope with these various challenges.[1]

4.1 Data Collection

Since the study context is local events, local event data was collected (1) from January 2017 to August 2018 (20 months) targeting 14 U.S. cities and (2) from June 2018 to August 2018 (3 months) targeting 29 U.S. cities (3 months). The 14 cities for the twenty months were selected based on the propensity of tech start-ups in those cities because the vibrant start-up culture of a city is related to diverse meetups in the region, not only stemming from organizations' efforts for their business success but also due to their high dependency on technological infrastructures for connecting with customers and collaborators [Risen, 2016]. The 29 cities for the three months were selected based on the large population rankings. The underlying assumption was that cities with large populations would have more events organized both by organizations and citizens.

The target information sources are (1) Meetup[2], a website that allows anybody to create and attend any kind of local event with "local groups" functionality, (2) Eventful[3], a website that allows users to create events and provides other available local events, and (3) Yelp[4], a website that is intended for providing local business information but also provides affordances to create local events that are organized in local venues. Each website provides Application Program Interfaces (APIs) from

[1] Using online data often comes with some biases as well, but it makes it possible to scale the study by virtue of a massive increase in the data volume and people's internet use in the 2010's.

[2] `https://meetup.com`

[3] `https://eventful.com`

[4] `https://yelp.com`

which researchers and developers can collect various kinds of data available on the website. Custom Python and PHP scripts are used to collect this data over time. If APIs allow, *cron job*, a Linux functionality that regularly runs particular scripts, is used to ensure data collection continuity.[5]

The event data collection stopped in early September 2018 because Meetup APIs stopped providing up-to-date events at the city level. Because it is a key of the study to quantify the fragmentation of local information, missing one information source is critical even if the datasets from the other sources are available. Due to this practical reason, the periods that data cover in this study are set to 20 months (for the 14 cities with high tech-startup activities) and 3 months (for the 29 cities with high populations), respectively. To triangulate the findings from these two datasets, the 14-city dataset for the same three months that the 29-city data is collected is also used.

Adding other event data sources would have been greatly beneficial for this study, but sampling and scaling information sources is one of the most difficult challenges because there are only a few information sources that are commonly available across many U.S. cities. From a quantification perspective, maintaining consistent information sources across all cities is essential in quantifying the fragmentation of local information and these three sources are chosen as the target information sources.[6]

[5] The scripts for collecting data are available at: `https://github.com/myeong/phd_dissertation/data_collection`

[6] Facebook Events was one of the strongest candidates due to the platform's popularity; however, Facebook stopped providing city-level event data in early 2018, which led me to omit this option.

Data Source		# of Events	Mean # per City	(SD)
Meetup	14 cities (20 months)	998,461	70,604	(42,953)
	29 cities (3 months)	196,554	6,778	(5,712)
Eventful	14 cities (20 months)	888,955	63,498	(61,225)
	29 cities (3 months)	420,709	14,507	(16,620)
Yelp	14 cities (20 months)	32,234	2,302	(1,951)
	29 cities (3 months)	7,260	250	(231)

Table 4.1: Descriptive statistics about the numbers of events in the collected datasets.

The number of events, average events per city, and standard deviation are presented in Table 4.1. The list of cities and the number of events for each city is available in Appendix A. These datasets provide comprehensive information about local events such as title, event venue address, geo-coordinates, and RSVP inforamtion. The list of the attributes available from each source is listed in Table 4.2. Figure 4.1 shows the geospatial distribution of Washington D.C.'s local events from the collected datasets in 2017.

Other datasets that are needed for this study are collected from the U.S. Census Bureau website.[7] Particularly, urban area-level geographical boundaries data in Shapefile, socio-economic status, socio-demographic information, and ethnoracial information estimation in 2017 are collected.[8] These datasets are downloaded in Comma-Separated Values (CSV) and Shapefiles formats. The API calls to the local event websites store the data in the JavaScript Object Notation (JSON) format. While these data formats are standardized, these datasets had to be prepro-

[7] U.S. Census: `https://factfinder.census.gov`

[8] As of May 2018, the demographic/economic data for 2018 is not available from the U.S. Census. Because the ACS data is created annually and does not change dramatically in a short period of time, 2017 data is reasonable to use for the purpose of this study.

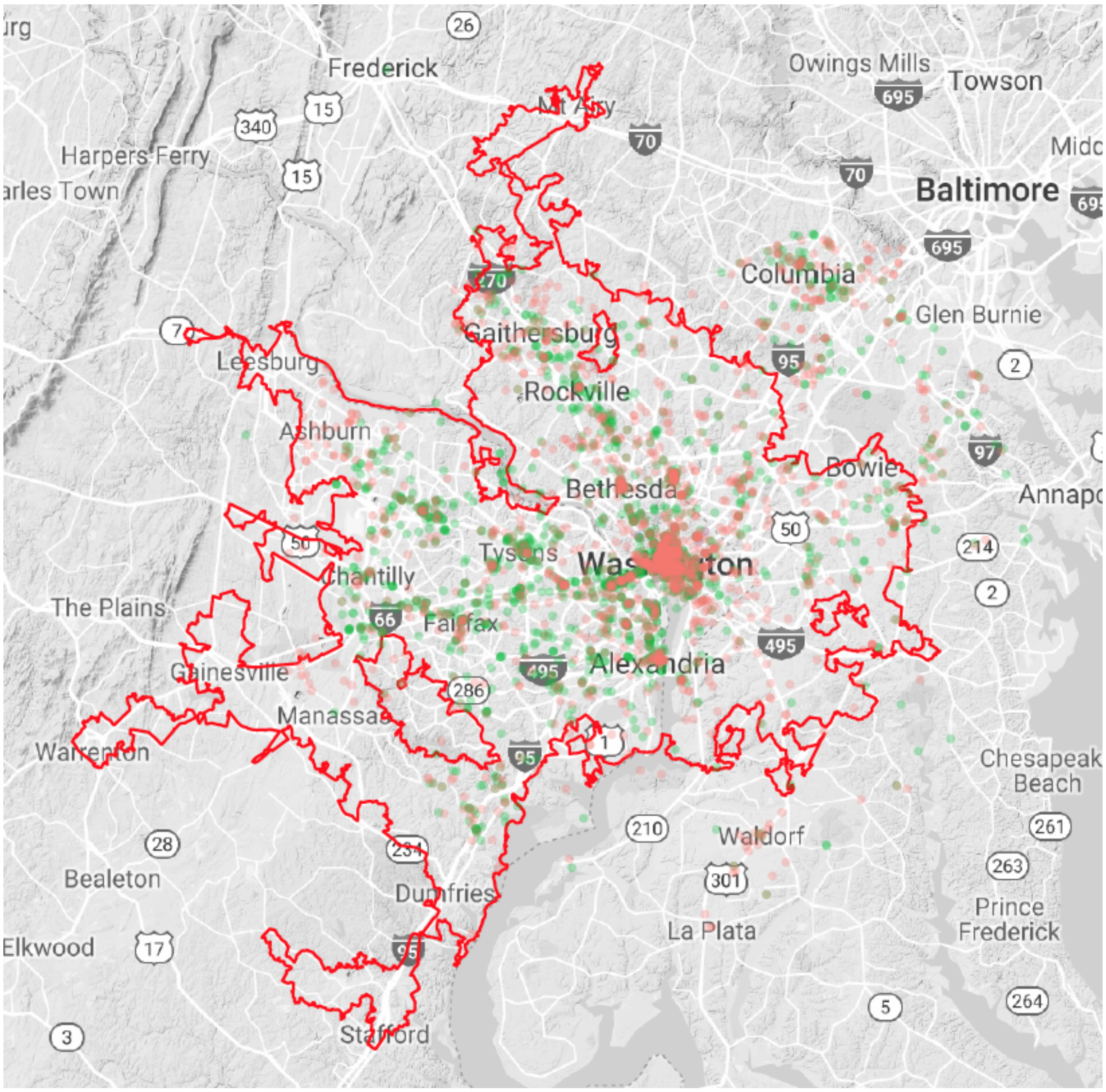

Figure 4.1: The locations of local events from different datasets in Washington D.C. for 2017: Meetup (green); Yelp (blue); and Eventful (red). Also, red boundaries show Washington D.C.'s urban area boundaries provided by U.S. Census Bureau.

cessed properly for further analysis either due to data format inconsistencies or some glitches embedded in them.

Urban area-level census regions (code 400) as the unit of analysis is reasonable given the datasets. Because local information is disseminated and used in a metropolitan area following their mobility range, using census tracts or municipal city boundaries as the unit of analysis would be too small to reflect information users' dynamics; conversely, using the state or census statistical metro/micropolitan boundaries are two broad. As can be seen in Figure 4.1, urban area boundaries do not follow municipal boundaries but cover urban regions with high population density well and are consistent with local event locations.[9]

While event data were collected for 29 cities for the 3-month dataset, the cities of Fort Worth and Dallas are one urban area in the census Shapefile (maybe because these two cities are connected with each other as a metropolitan area). Due to this reason, we use 28 urban areas for the following analysis of the 3-month dataset.

4.2 Preliminary Data Processing

As shown in Table 4.2, there are some inconsistencies between the attributes available in the datasets. For example, Meetup data provides *start_time* as an attribute but does not provide *end_time*; instead, it provides the *duration* field so to calculate the ending time of each event. If *duration* is not specified in a Meetup record, the API documentation states that the default duration of an event is three

[9] The map shows that some events are not included in the urban area boundaries. However, the majority of events is included in these boundaries across different cities.

Attribute	Description	Meetup	Eventful	Yelp
Event ID	The unique ID of the event	1	1	1
Event title	The title of the event	1	1	1
Description	The description about the event	1	1	1
Event URL	Domain address for the event	1	1	1
RSVP	The number of RSVPs	1	0	1
Start time	The start time of the event	1	1	1
End time	The end time of the event	0	1	1
UTC offset	Local time difference compared to UTC	1	0	0
Time zone	Local time zone of the event location	0	1	0
Duration	The duration of the event	1	0	0
All day	Whether this event lasts all day	0	1	0
Venue ID	The unique ID of the event venue	1	0	1
Venue name	The name of the event venue	1	1	0
Venue address	Street address of the venue	1	1	1
State	The state name of the venue	1	1	1
City	The city name of the venue	1	1	1
Zip	The zip code of the venue	1	1	1
longitude/latitude	The geo-coordinates of the venue	1	1	1
Geocoding type	The precision of geocoding	0	1	0
Venue repinned	Manually geo-located?	1	0	0
How to find us	Descriptions about how to find the event venue	1	0	0
Created	Event posting date/time	1	1	0
Category	The category of the event (topic)	0	1	1
Group name	The name of the organizing group	1	0	0
Join-mode	Membership open to the public?	1	0	0

Table 4.2: Available attributes from each of the raw dataset. Only key attributes that are relevant to this study are presented. Even if a field exists in a dataset, if there is no value in the column, the availability of the attribute is 0 in the table.

hours.[10] Yelp and Eventful sometimes do not include *end_time* or *duration*. In this case, the end time is set to the end of the day (e.g., the end time of an event that begins at 6 p.m. without any duration information is set to 11:59 p.m. on the same day).

In addition to the start/end time generation, UTC offset is also taken into account in adjusting the times. Meetup and Eventful provide UTC time as the reference timezone with an additional attribute, *utc_offset* or *time_zone*. Because the events in the data are from cities in different time zones, the offset information is used to adjust all the times in the datasets so each event can have local timestamps in the data.

Text data is also cleaned up. Because each of the raw datasets is in different styles of texts such as inclusion of HTML tags, special characters, and text length, it is necessary to process them so to make them consistent. All the tags and special characters are removed from *event_title* and *event_description*. Also, all the words in these fields are changed to lower case and stemmed for text processing. Multiple white spaces between words are reduced to one space. After conducting this preliminary data processing against all the three event datasets, the data is saved as CSV files.[11]

[10] Meetup API documentation accessed as of May 2019: `https://www.meetup.com/meetup_api/docs/2/open_events/`

[11] The data cleaning scripts are available at: `https://github.com/myeong/phd_dissertation/data_cleaning`

4.3 Event Data Disambiguation

Because a core part of this study is to measure the fragmentation of local information, it is essential to quantify the extent to which local information is scattered or dispersed across different information sources. The first step for this process is to identify duplicate events between different information sources. However, there is no straightforward way to capture duplicate events because there is no shared key or ID between two or more websites. Furthermore, event titles and descriptions from different sources are often inconsistent with each other, even if they are physically the same event (maybe due to ad-hoc provision of local information by different actors). This suggests that event data needs to be disambiguated with multi-step, computational treatments rather than using a naïve text matching.

Data disambiguation techniques have been developed extensively in the field of information science in which many of these studies aimed to detect academic communities through co-authorship networks of scientific articles [Kim and Diesner, 2016, Yu et al., 2017]. While co-authorship disambiguation techniques often make use of relatively well-curated information such as author names and their affiliations (which is already a hard problem due to limited information in the data), the errors in the event data (e.g., inaccurate location information) and inconsistent titles/descriptions of some physically-same events make it further complicated to directly use the existing methods.

In this regard, this study disambiguates pieces of local event information from different sources by making use of machine learning techniques along with other

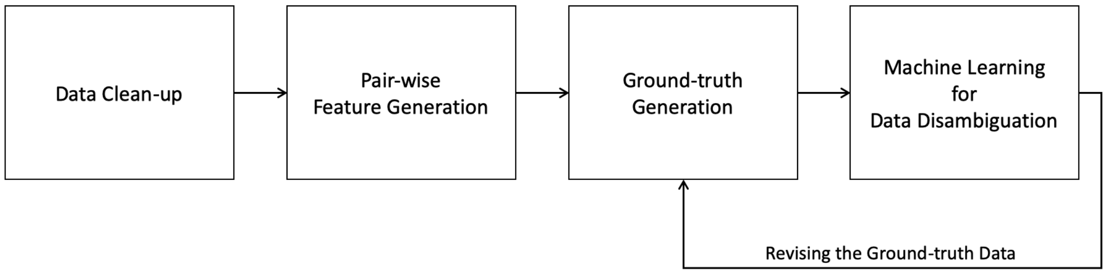

Figure 4.2: Overall process used to disambiguate local event information from different sources.

computational methods. The overall process is briefly depicted in Figure 4.2. At the core of this approach, as with other prediction studies, generating and engineering features for machine learning models are the most important and fundamental step.

4.3.1 Feature Engineering

Using the cleaned datasets, features for machine learning models are generated between each pair of events (thus, pair-wise features). Because the purpose of using machine learning is to scale the decisions on whether two different events are the same event or not (i.e., *match/non-match*), generated features signify relational, pair-wise characteristics that indicate the similarity between two events. These features consist of semantic, temporal, and physical similarity indicators. The final list of the features is listed in Table 4.3.

The start times in the events data are relatively accurate across different datasets. Although the availability of events' end times are often inconsistent and unpredictable, start time is one of the powerful predictors in identifying match/non-match between two events. Time overlap hours and overlap rates provide other use-

Feature	Description
Start time difference	The start time difference between two events in hours.
Time overlap hours	The overlapping hours between the two events' time periods.
Time overlap rate	The extent to which two events' time periods overlap given the total time period of the two events. $$overlap\_rate = \frac{t_{overlap}}{t_{max} - t_{min}}$$
Title similarity	The Jaccard similarity between two events' names.
Description similarity	The Jaccard similarity between two events' descriptions.
Physical distance	Physical distance between two events' locations (km).
Geocoding type	The lowest resolution of geocoding type between two events.

Table 4.3: Pair-wise features generated for machine learning models. All the features are numeric values with floating points, except for *geocoding_type* (which is a categorical feature).

ful information, especially for records that have end time information. Even though end times are unclear for certain events, automatically-generated end times provide reasonable estimates in finding match/non-match.[12]

Jaccard similarity is used for measuring the similarity based on event titles/descriptions' bag-of-words models. Other similarity/distance measures such as Euclidean distance and cosine similarity could be used, but Jaccard similarity is selected based on previous work on the benchmark of document distance measures, which showed superior performance of Jaccard similarity in sparse document clustering tasks [Strehl et al., 2000].

The distance between two event locations is also a powerful indicator to pre-

[12]Given the fact that most of the non-match events are distinguished based on event dates rather than time overlaps, the uncertainty in the end times within a day is justifiable.

dict match/non-match. However, a complication in the geo-coordinates data of the Meetup datasets is that some Meetup records do not contain location information at all or present inconsistency between geo-coordinates (i.e., *longitude* and *latitude*) and the physical address of their event locations (e.g., *venue_address* and *city*). These inconsistencies/errors stem from the nature of user-generated location data on the web platform that provides high flexibility in specifying event locations.

The data quality of location information in the Eventful and Yelp datasets, in the meantime, is slightly different from that of Meetup. Through qualitative examinations, it is found that Yelp data has high-quality location information for each event. There is no inconsistency found between geo-coordinates and physical addresses in random samples of Yelp events. Although Eventful data has some issues in the location data, similar to Meetup's, Eventful data provides a useful attribute called *geocoding_type*.

Because Eventful data has been curated not only by the users of the platform, but also through the company's own data migration practices, each event record is tagged with one of the geocoding types: *place-level*, *zipcode-level*. and *city-level*.[13] Place-level geocoding provides the precise geo-coordinates of an event; zipcode-level geocoding provides less-precise information about longitude/latitudes but in a zip code-level accuracy; and city-level geocoding indicates the inaccurate location information of the event data within a city. Even though there are some errors found in the geo-coordinates of Eventful events, considering the geocoding classifications

[13] Actual values in the raw data that define these three levels of geocoding types are different, but for the sake of readers' easy understanding, these three classification names are used throughout this chapter.

makes it possible to identify whether their location information is precise or not.

Using the geocoding type attribute, it is possible to complement the errors in Meetup's geo-coordinates because there is no accuracy indicator for the location data and the number of errors is not negligible. In this regard, Meetup's location data is examined and improved through a series of data processing. Particularly, each record of the Meetup location data is examined automatically using scripts and, if necessary, is tagged with one of the geocoding schemes following those of Eventful. The automatic adjustment process for the geo-coordinates is as follows. In each step, the number of processed event records is presented for the Washington D.C. data in 2017 to provide a sense of how many locations are corrected and tagged throughout the process (the total number of events for Washington D.C. in 2017 is 73,759):

Step 1: Making use of "venue_repinned." One observation is that if the *venue_repinned* attribute of a Meetup event is $TRUE$, meaning the event organizer manually picked a particular location on the map for the event, its geo-coordinates is accurate; thus, the records with $TRUE$ for the *venue_repinned* attribute are tagged "place-level" for the *geocoding_type* attribute. For Washington D.C.'s 2017 data, for example, 342 events are tagged "place-level" for their locations with precise geo-coordinates.

Step 2: Geocoding for unregistered venues. Then, the event venues are examined to see if there are any anomalies. When the venue ID of a Meetup event (*venue_id*) is not available, it means this event does not make use of registered venue data from Meetup's venue database; rather, the venue information is manually entered by the organizer in the *how_to_find_us* field *or* not available at all. If both *venue_id* and *how_to_find_us* attributes are not available for an event, this event is tagged with "city-level" and the center of the city is used for the geo-coordinates because venue information is not available at all (for Washington D.C. data, 13,366 events are tagged "city-level" for their locations).

If *venue_id* is not available but *how_to_find_us* exists, it is possible that the *how_to_find_us* attribute contains one of the following: physical address, geo-coordinates, a URL of further information, email address of the organizer, or the phone number of the organizer. Regular expressions are used to detect the form of the information in this attribute, and location information is extracted from the *how_to_find_us* field when relevant. For example, if the pattern matches longitude/latitude, the value is extracted and copied to the geo-coordinates attributes; then, this event is tagged "place-level" for the geocoding type (for Washington D.C.'s 2017 data, four events have geo-coordinates in the *how_to_find_us* field). If the pattern matched email address or URL, these records are ignored from the processing.

If any of the pre-defined patterns are not relevant to the *how_to_find_us* field, the text is assumed as a physical address and geocoded using the Google Maps Geocoding API.[14] For Washington D.C. data, 307 events' locations are tagged with "place-level" with precise geo-coordinates in this step. When the Geocoding API returns an error or *null* value, the event record is tagged with "city-level" for the *geocoding_type* attribute. Also, when the geo-coordinates returned from the API

[14] `https://developers.google.com/maps/documentation/geocoding/start`

is more than 25 miles from the city center, the record is tagged with "city-level" because it is highly possible that this geocoding value is wrong (mostly because the text in the *how_to_find_us* field does not correctly present a physical address).

Step 3: Geocoding for registered online venues and TBD's. It is possible that registered venues on Meetup are actually not physical locations but online addresses, especially for the events that happen through web interfaces (e.g., webinars, web-casts, radio events, etc.). Also, it is possible that *venue_address* shows "TBD" and provides a zip code- or city-level location only. Regular expressions are used to detect online events, which are tagged with "city-level" (for Washington D.C. data, 288 records are detected). If place-level location data is not available or tagged "TBD," other location-related fields such as *city* and *zipcode* attributes are examined to check the granularity of the location information. If *venue_address* does not exist, the finest resolution of location information is geocoded using Google Geocoding API and tagged accordingly. For example, if a record has *zipcode* but does not have a physical address, it is geocoded using the center of the zip code region and tagged with "zipcode-level" for the *geocoding_type* attribute. Through this process, for Washington D.C. data in 2017, 288 venues were updated along with the *geocoding_type* attribute.

Step 4: Geocoding for registered venues with no geo-coordinates. There are registered venues that contain physical addresses but without geo-coordinates. In this case, the physical address is geocoded using the Google Geocoding API. Sometimes, a physical address does not provide the street-level address; in such cases, similar to Step 3, it is geocoded based on the finest granularity of location available in the data. For a sanity check, the distance between the geo-coordinates from API and the center of the city is calculated for each record. If this distance is longer than 25 miles, the geocoding result is assumed to be wrong, maybe due to incomplete address information, and tagged "city-level." Depending on the precision of the geocoding, each record is tagged with one of the three geocoding classifications. In this step, 421 registered venues' geo-coordinates were updated for Washington D.C.'s 2017 data, which resulted in updating 2,459 events' location information.

Step 5: Random checks for the rest of the location data. The rest of the event data has both physical addresses and geo-coordinates. Since it is possible that there are still some errors in the location data (e.g., it is possible that the geo-coordinates of an event location do not point to the physical address associated with it), 2,000 events are randomly selected from the events that do not have any issues in the location data and their physical addresses are geocoded using the same API. Then, the distances between the geocoded coordinates and original geo-coordinates from the data are calculated to see if they point to the same location. This check shows that about 6.5% of the random samples present more than 1.6km difference (1 mile) between the geocoded coordinates and original geo-coordinates. Although this is not ideal, this error rate is reasonable for further analysis when all the error corrections until Step 4 are considered. For example, it comes down to less than 5% error rate in the location data in the case of the 2017 Washington D.C. data. Hence, all the event datasets are processed using this five-step process to correct the location data and tag them with one of the three geocoding types. The overall process is debriefed in Algorithm 1.[15]

4.3.2 Machine Learning and Performance Assessment

Generating ground-truth data. After generating and engineering features, ground-truth data is generated by manually coding matches/non-matches for certain pairs of event records from different sources. Particularly, all the pairs of Eventful and Yelp events are generated along with all the pair-wise features. Then, pairs with no overlap in their time periods (i.e., $time\_overlap\_hours = 0$) are removed from the pair dataset because there is no possibility that they are the same event. These pairs of events are sorted in a descending order based on similarity scores. From the top records that show the highest similarities in their features, I manually compare each pair of events by visiting the corresponding event's web page to check whether each pair is physically the same event.

Initially, the match/non-match is coded in two ways where one is conservative coding and the other is flexible coding. The conservative codings assess the match/non-match of two events based on their physical location and organizer; in the meantime, the flexible codings focus only on their relations to a physical event. For example, Science March was a nation-wide event in 2017 in Washington D.C. and many different Meetup groups created Meetup events online for the march. While many groups gathered in different locations, they physically participated in the same event. When organizers are different for Science March (so event de-

[15] Scripts for the location data correction process are available at: `https://github.com/myeong/phd_dissertation/feature_generation`

Algorithm 1: The data correction and tagging process for the locations of Meetup events.

```
Input: Given a Meetup event record E
if E.venue_repinned is TRUE then
    geocoding_type = "place-level" /* E.coordinates are correct. */
else if E.venue_id is NULL and E.how_to_find_us is NULL then
    if E.zipcode is available then
        E.geocoding_type = "zipcode-level"
        E.coordinates = Zip_Centroid.coordinates
    else
        E.geocoding_type = "city-level"
        E.coordinates = City_Centroid.coordinates
else if E.venue_id is NULL and E.how_to_find_us exists then
    if E.how_to_find_us matches to non-venue patterns (e.g., email, website...)
     then
        E.geocoding_type = "city-level"
        E.coordinates = City_Centroid.coordinates
    else if E.how_to_find_us matches to lon/lat patterns then
        E.geocoding_type = "place-level"
        E.coordinates = E.how_to_find_us
    else
        Assert (C = geocoding(E.how_to_find_us))
        if C.Error or C is NULL or distance_check(C) fails then
            E.geocoding_type = "city-level"
            E.coordinates = City_Centroid.coordinates
        else
            E.geocoding_type = "place-level"
            E.coordinates = C
else
    if E.coordinates not exists then
        E.coordinates = geocode(E.address)
        E.geocoding_type = "place-level" /* If distance_check passes */
    else
        distance_check(E.coordinates, City_Centroid.coordiantes)
```

scriptions/locations from websites are also different), conservative coding tags it as "non-match" because their gathering locations and organizers are different; conversely, flexible coding tags it as "match" because the physical event that they actually attended is the same regardless of the organizers.

Through qualitative examinations and machine learning (ML) performance tests, the flexible codings are decided as the main object variable because theoretically it makes more sense to focus on physical events rather than people who create online events and, on a more practical level, flexible codings yield a better consistency in the ML tests (maybe because the organizer inconsistency between different websites is eliminated in the flexible codings).[16]

After manually coding about 500 pairs of events in the sorted dataset, the pairs started saturating to "non-match," so the manual coding stopped at the 557th record of the pairs.[17] With this manual coding results as the ground-truth data, the performances of different machine learning models are tested by randomly training a proportion of the dataset and predicting the remainders using the trained model.

ML models and baseline performance. ML models used are Random Forests, an ensemble learning method [Breiman, 2001], Support-Vector Machine (SVM), a supervised learning method based on hyperplane segmentations [Cortes and Vapnik, 1995], and Decision Tree, a tree-based decision-support tool for classification tasks [Breiman, 2017].[18] Identifying whether a pair of events is a match or

[16]In testing ML-based predictions, conservative codings yielded at best 0.90 in the F_1 score, while flexible codings yielded 0.96 in its best performed model.

[17]The number of pairs per month is very large (e.g., around 10K records for a month in Washington D.C. between Yelp and Eventful), but due to a lack of overlapping events between different websites, coding about 300 events already started to not show matches.

[18]Deep learning was also tested but its performance fluctuated too much due to the small

a non-match is a 2-bin classification problem. Naïve prediction without using any algorithms could be 50% guess on match/non-match for each pair of events, but due to the nature of data where there is only a small number of matches, 50% random guess is not appropriate for a baseline performance test. In other words, if someone guesses all the pairs as "non-match," the agreement of the prediction would become very good.

For this reason, predictions with random features are used in providing the baseline performance [Smith-Clarke and Capra, 2016]. Specifically, 10,000 random numbers that follow the normal distributions of actual similarity measures are created. Then, 557 random samples from the generated numbers are used as the features for predicting match/non-match. To ensure the robustness of the model, the performance is measured based on the average F_1 score of 100 independent predictions. F_1 score is used as the ML model performance metric because it takes both precision and recall into account, which is more useful and precise than using accuracy when the number of matches is way smaller than that of non-matches.

$$F_1 = 2 * \frac{precision * recall}{precision + recall} \tag{4.1}$$

Similar way to baseline performance testing, machine learning models are tested using generated features in Table 4.3, and their performances are measured using F_1 scores based on the average of 100 test results. For each test of the models, the proportions of the training set range from 50% to 90% to show the robustness

number of training sets. This led me to omit the DL models from candidates.

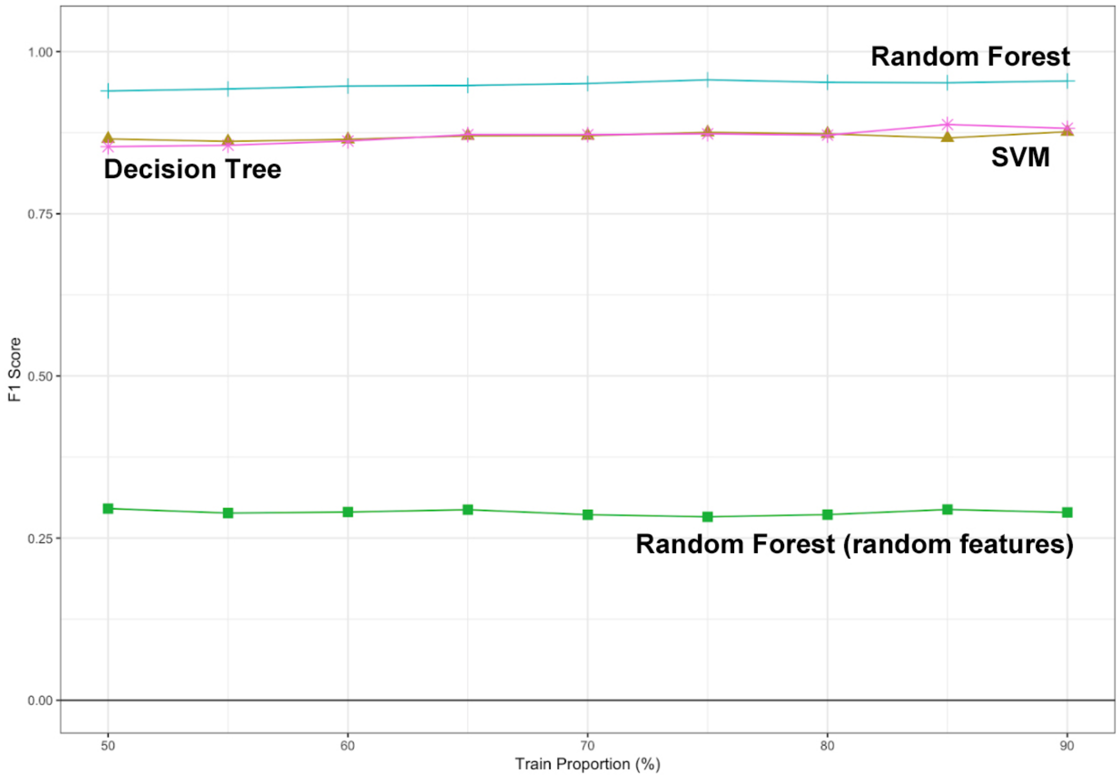

Figure 4.3: F_1 scores of different machine learning models. Each score is the average of 100 independent predictions.

of the model. The performances of the machine learning models within the 557 event-pairs data are shown in Figure 4.3.

The F_1 score of the baseline model is about 0.290 when Random Forests is used with random features. F_1 scores of the other ML models for the baseline tests are not shown in the graph because there is not true-positives, which make precision undefined (so F_1 is undefined). Random Forests presents the best performance in its average F_1 score, 0.955, when 90% of the records are trained by the model. SVM and Decision Tree show good performances as well, but F_1 scores are 0.876 and 0.882, respectively.

Amazon M-Turk as an alternative method? Crowdsourcing could also be

N=557	Actual True	Actual False
Predicted True	147	45
Predicted False	18	347

F_1= 0.82, *Recall*=0.89, *Precision*=0.77

Table 4.4: The performance of M-Turkers against manually-coded ground-truths for disambiguating event data.

a powerful method to scale up data disambiguation. To check the possibility, the 557 records that I manually coded were submitted to Amazon M-Turk and three different Turkers rated each pair of events whether it was a match or non-match. When two or more Turkers agreed upon the match/non-match, it was tagged as "match" and otherwise as "non-match." The performance of Amazon M-Turk against my own manual coding is shown in Table 4.4 as a confusion matrix. The F_1 score of Turkers' codings is 0.824, which is lower than the ML models' performances.

For a further evaluation of the usefulness of M-Turk, assuming Turkers' answers for match/non-match values are ground-truth, the three ML models are used to train the proportions of the 557 records and tested against the rest of the records to see if M-turk could be useful in generating ground-truth data for ML models. Among all the prediction tasks, Random Forests performs the best when using the M-turk results as the ground-truth data (F_1=0.752). This suggests that Amazon M-Turk is not powerful enough as the ground-truth data as well, compared to manual codings, in predicting match/non-match for each pair of events. Therefore, the disambiguation process relies on the manual codings of match/non-match and ML models trained based on them.

ML model adjustments. These performance tests allowed for further ex-

RF	Actual True	Actual False
Predicted True	165	104
Predicted False	0	56,623

F_1= 0.76, *Recall*=1.00, *Precision*=0.61

SVM	Actual True	Actual False
Predicted True	141	9
Predicted False	24	56,718

F_1= 0.90, *Recall*=0.85, *Precision*=0.94

D/Tree	Actual True	Actual False
Predicted True	146	86
Predicted False	19	56,641

F_1= 0.74, *Recall*=0.88, *Precision*=0.63

Table 4.5: Confusion matrices of ML prediction results for testing the scalability of ML models for Washington D.C.'s 2017 data (N=56,892).

amination of the trained models to scale the prediction tasks. Because the manual coding stopped at the 557th record in the descending-ordered list of event pairs and the rest of the records is assumed as "non-matches," all of the uncoded records for the Washington D.C. data in 2017 are tagged with "non-match" and the aggregated records that include both coded and assumed values are used as ground-truth for evaluating the ML models (totally, N=56,892). The confusion matrices of these tests are shown in Table 4.5.

The results show that the best F_1 score is based on the SVM model, which is different from the result when predicting only the manually-coded records. However, there are a couple of aspects that need to be noticed. First of all, recall is more important than precision in this kind of dataset where the number of matches is very small. ML models need to be able to find matched pairs of events rather than conservatively predicting the target variable. In other words, even if the model has to sacrifice precision, it is important to find all the true matches. Because of this reason, finding true-positives and false-positives is more meaningful than increasing the number of false-negatives. From this perspective, Random Forests still provides more meaningful results than other models.

To ensure the performance of the models, all the false-positives and false-negatives from all three models are manually examined to see if there are mis-classified records. Surprisingly, 17 records among the false-positives of both Random Forests and Decision Tree models, respectively, are identified as true-positives. Conversely, none of the false-negatives and false-positives from the results of the SVM model is identified as true-positives. This suggests that tree-based ML models are superior in increasing recall and finding true matches in the given dataset. Between the Random Forests and Decision Tree, Random Forests is still better in precision.

After adjusting the ground-truth data by incorporating the result of the manual examinations for the 109 falsely-predicted values which include the 17 additional true-positives, in total 666 records are used for the final ML model construction. Using this new ML model that is trained based on the 666 ground-truth records, the prediction power increases significantly as shown in the confusion matrix (Table 4.6). This model is used for finding matches and non-matches for the entire event pair dataset.[19]

[19] Scripts for the machine learning process are available at: `https://github.com/myeong/phd_dissertation/events_disambiguation`

N=182	Actual True	Actual False
Predicted True	182	1
Predicted False	0	56,709

F_1= 0.99, *Recall*=1.00, *Precision*=0.99

Table 4.6: The performance of the enhanced Random Forests model for the Washington D.C.'s 2017 dataset.

4.3.3 Statistical Imputation

Using the ML model constructed, all the target cities' events are disambiguated for each month. Then, the changes of event volumes and duplicates between each pair of the three sources are generated to check the data patterns. For example, Figure 4.4 depicts an example of month-to-month event volume changes. In the figure, it is possible to see that the number of Meetup events decreases significantly in October 2017 and increases back to normal in November 2017, while the other datasets show smooth changes over time. This pattern is observed in all the cities. After some qualitative examinations, it is found that there is missing Meetup data for 14 days in October 2017 due to an API malfunction. Because missing data significantly affects the key variables, the generated data could not be used directly for further analysis. To deal with this challenge, statistical imputations are used to fill the data gaps.

For the number of Meetup events for the missing period, it is possible to prorate the number of events using the rest of the month's events because 17 days' Meetup data for the rest of October 2017 is available. The number of Meetup events in October 2017 for each city, thus, is determined by prorating the numbers for the

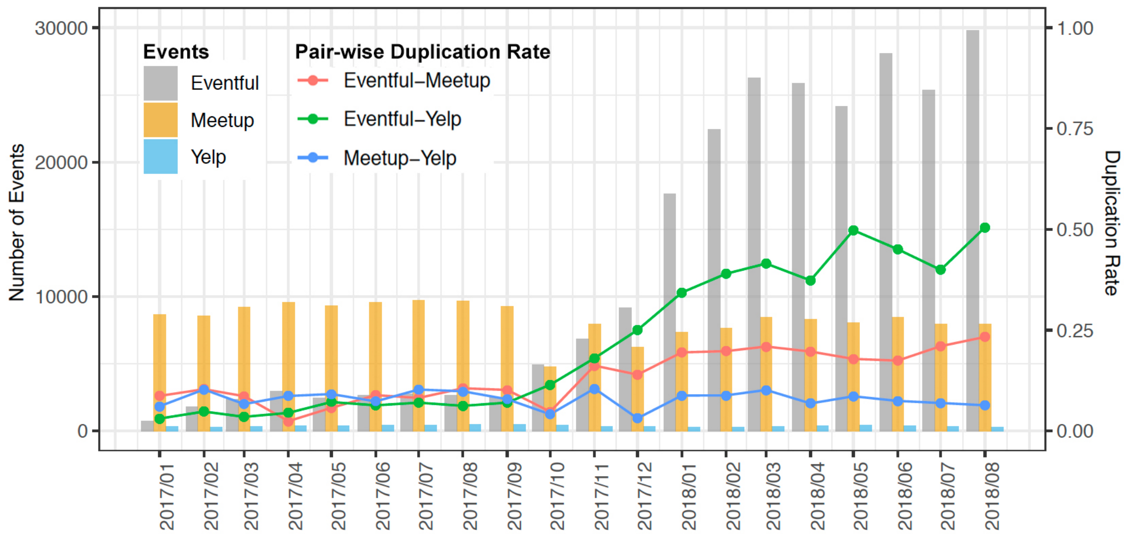

Figure 4.4: An example of event volume and duplicate changes over time for New York City, NY. The number of events in October 2017 is odd due to missing data for two weeks.

17-day data.

$$N_{October\_events} = \frac{31}{17} * N_{17day\_events} \tag{4.2}$$

However, the number of duplicates between different sources cannot be pro-rated because event duplication is not a function of the number of Meetup records. Instead, multi-variate regressions are used to predict the numbers of duplicates. The underlying presumption is that the number of duplicates is determined by the numbers of Meetup, Yelp, and Eventful records in addition to the effects of time and space. Therefore, the multi-variate regression model for predicting the number of duplicates, $N_{dup(t,i)}$, in month t and city i is as follows:

$$N_{dup(t,i)} = \beta_0 + \beta_1 N_{meetup(t,i)} + \beta_2 N_{eventful(t,i)} + \beta_3 N_{yelp(t,i)} + \beta_4 D_t + \beta_5 C_i + \varepsilon_{(t,i)} \tag{4.3}$$

where D_t is time dummy variable for time t (year/month) and C_i is city dummy

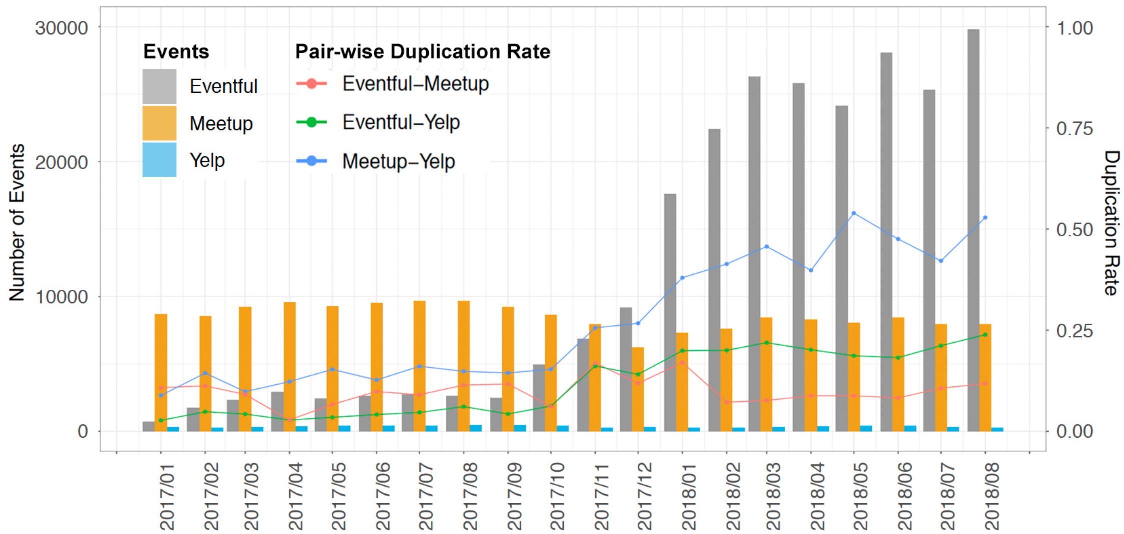

Figure 4.5: An example of event volume and duplicate changes over time for New York City, NY, after imputing the number of events and duplicates in October 2017.

variable for city i. For the pair-wise duplicates, relevant variables are used accordingly. For example, to predict the number of duplicates between Meetup and Yelp, the β_2 term is omitted from the regression model because the number of Eventful records has nothing to do with the overlap between Meetup and Yelp. An example of the adjusted data for New York City is presented in Figure 4.5 in which it is possible to see that the changes of Meetup volume and duplicates have smoothened after the statistical imputations. These adjusted numbers are used for further analysis in calculating the fragmentation of local information.[20]

[20]Scripts for the imputation process are available at: `https://github.com/myeong/phd_dissertation/events_disambiguation`

4.4 Computational Modeling of the Fragmentation of Local Information

Based on the number of organized events in each information source (i.e., N_{meetup}, N_{yelp}, and $N_{eventful}$), their pair-wise duplicates (i.e., $N_{m \cap y}$, $N_{m \cap e}$, and $N_{e \cap y}$), and the number of overlapping events across all the three sources (i.e., $N_{m \cap y \cap e}$) for each month and city, it is possible to capture the overall duplication rate changes over time as can be seen in Figure 4.5. However, the concept of "fragmentation" is not established or widely-used in the local information context. In other words, because this is the first study that measures the fragmentation of local information, deciding an appropriate measure is not a trivial problem; rather, it is necessary to choose or develop the measure in a systematic way based on conceptual and methodological justifications.

What this study aims to measure with regard to the concept of information deserts is *how much the pieces of local information in a city are dispersed and segregated across different information sources*. This characterization is very similar to the concept of *inequality* and *diversity* while their meanings and nuances are very different. For example, Gini index, a representative economic inequality measure, operationalizes the extent to which income and wealth are dispersed across members of a community based on a distribution of them [Gini, 1936], which is conceptually similar to what the fragmentation of local information aims to quantify. Gini index becomes 0 when all the people in a community have equal amounts of wealth and

1 if one person has all the wealth while the others have none. Firebaugh [1999] explains that Gini index is a special form of inequality entropy where its generalized form can be expressed as follows:

$$I = \sum_j p_j f(r_j) \tag{4.4}$$

where $r_j = X_i/\overline{X}$, the income ratio for the jth unit; $p_j = n_j/N$, the population share of the jth group, with a constraint, $\Sigma_j f(r_j) = 0$ when r_j=1 for all j. This general form is applicable to other well-known diversity measures such as *Shannon entropy* [Shannon, 1948] and *Simpson index* (a.k.a., *HHI index* [Hirschman, 1964]) that have been frequently used to measure diversity when some elements are assigned to different classes, particularly in the studies of ecology, economics, and social sciences [Keylock, 2005, Hu et al., 2019, Butler, 2001, Hristova et al., 2018]. Their forms are instances of the generalized inequality entropy formula and shown as follows:

$$I_{Shannon} = -\sum_j p_j \log p_j \tag{4.5}$$

$$I_{Simpson} = \sum_{j=1}^{R} p_j^2 \tag{4.6}$$

where R is the number of groups. Although the general form of inequality/diversity and its variations such as Equations 4.5 and 4.6 are conceptually consistent with the meaning of the fragmentation of local information, there is a key difference between these two notions. Most of the inequality entropy measures quantify the dispersion of population shares when they are distributed across multiple groups or categories.

However, inequality entropy does not take the overlapping shares between groups into account. One of the key characteristics of the fragmentation of local information is that pieces of event information are not only dispersed across different information sources, but also duplicated between multiple sources (i.e., overlaps). Both of these two dimensions need to be considered in the fragmentation measure to precisely operationalize the original concept.

Types of diversity measures. With respect to the operationalization of these two different dimensions of diversity or inequality, the field of ecology has a long tradition in using and developing such diversity measures to quantify species distributions in different habitats [Jost, 2007, Peet, 1974]. Part of the key ecological concepts such as *richness*, *evenness*, and *heterogeneity* have been adapted to social sciences as well to explain the effects of social- and organizational-ecology on organizations' survival rates [Hannan and Freeman, 1993, Wang et al., 2016b]. While social sciences scholars have adapted part of the diversity measures from the field of ecology, measuring the fragmentation of local information can further benefit from the systematic conceptualizations and various measures from the ecology literature.

Ecology scholars conceptualize and characterize diversity in three dimensions: *alpha*, *beta*, and *gamma* diversities [Jost, 2007]. Alpha diversity quantifies the diversity of species within a habitat or region. These measures are relatively well known due to the needs for quantifying diversity in various communities; these include, but are not limited to, Simpson index, Shannon entropy, and Rao-Stirling index [Stirling, 2007, Rao, 1982]. These measures, as discussed, do not take species overlap between different habitats into account. Meanwhile, beta diversity conceptualizes

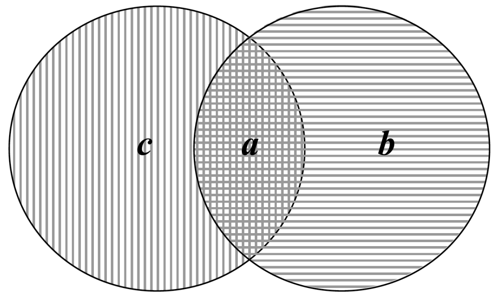

Figure 4.6: A baseline Venn diagram for ternary plots of beta diversity measrues [Koleff et al., 2003].

the between-communities or between-habitats diversity to describe the extent to which species are dispersed across different regions as well as how much they overlap with each other in the ecology [Koleff et al., 2003]. Because species can appear in many different habitats, species overlaps are an important part of beta diversity. Finally, gamma diversity is a combination of alpha and beta diversities and often expressed as either the sum or multiplication of them.

Among these dimensions of diversity, beta diversity properly conceptualizes the fragmentation of local information by taking both species overlap and dispersion (or balance/evenness) into account. In the context of event information, a physical event plays a role as a species (so, a disambiguated event can appear in two or more websites, which is an event overlap).

Beta diversity measures. While Whittaker [1960]'s beta diversity measure (β_w) has been most widely-used, Koleff et al. [2003] provides a thorough and extensive review of different beta diversity measures developed in the field of ecology. Because both evenness (or balance) and overlap are important in beta diversity, Koleff et al. [2003] make use of ternary plots to visualize the characteristics of dif-

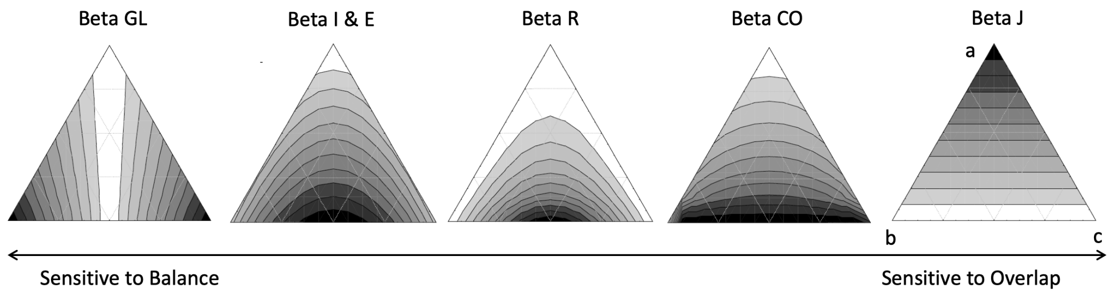

Figure 4.7: Ternary plots for beta diversity measures when plotted based on the Venn diagram in Figure 4.6 [Koleff et al., 2003]. The darker a part of the plot is, the higher the diversity is.

ferent beta diversity measures. Particularly, ternary plots are presented based on a basic, two-set Venn diagram for various beta diversity measures. The Venn diagram and ternary plots of some beta diversity measures that Koleff et al. [2003] present in their paper are shown in Figure 4.6 and Figure 4.7.[21] The formulas for the diversity measures in Figure 4.7 regarding the Venn diagram are as follows:

$$\beta_{GL} = \frac{2|b-c|}{2a+b+c} \qquad \text{[Lennon et al., 2001]} \tag{4.7}$$

$$\begin{aligned} \beta_I = \log(2a+b+c) - (\tfrac{1}{2a+b+c} 2a \log 2) - \\ (\tfrac{1}{2a+b+c}((a+b)\log(a+b) + (a+c)\log(a+c)) \quad \text{[Routledge, 1977]} \end{aligned} \tag{4.8}$$

$$\beta_E = exp(\beta_I) - 1 \qquad \text{[Routledge, 1977]} \tag{4.9}$$

$$\beta_R = \frac{(a+b+c)^2}{(a+b+c)^2 - 2bc} - 1 \qquad \text{[Routledge, 1977]} \tag{4.10}$$

$$\beta_{CO} = 1 - \frac{(a(2a+b+c))}{2(a+b)(a+c)} \qquad \text{[Cody, 1993]} \tag{4.11}$$

$$\beta_J = \frac{a}{a+b+c} \qquad \text{[Jaccard, 1912]} \tag{4.12}$$

[21] Some measures are opposite to others, which could be reversed when needed.

In Figure 4.6, a is the intersection between the two sets, and b and c represent the pure part of each set except the overlapping part, a. In the ternary plots, the center peak of the plot is a and both sides of the bottom edge are b and c, respectively. The ternary plots basically show how each diversity score changes when the combination of a, b, and c changes on the plane. For example, assuming each variable can range from 0 to 100, if a spot is located at the center peak of the ternary plot, a is 100 and both b and c are 0. If a spot is located at the center of the plot's bottom line, a is 0 and b and c are 50, respectively.

Based on the plots, it is possible to observe that beta diversity measures quantify the combination of balance and overlap between the two sets while putting weights differently on them. As indicated in Figure 4.7, β_{GL} puts most of the weight on balance while β_J, an equivalent to Jaccard similarity, measures only overlap between the two sets. The other measures range between these two extremes while putting weights on balance and overlap differently. β_I and β_E are relatively more sensitive to the balance of volumes than β_R or β_{CO}. β_{CO} is less sensitive to balance and put more weight on overlap but still varies slightly when balance changes. This indicates that beta diversity measure needs to be selected depending on variations of balance and overlap, which are contingent on context, conceptualizations of phenomena, and their analysis. Because it is not yet clear how the balance and overlap of pieces of event information vary within the local event datasets, it is necessary to descriptively analyze the data with respect to beta diversity measures.

For this reason, all these measures are used for subsequent analyses to assess the internal and external validity of the fragmentation measures in the context of local event information. Even though it is challenging to finalize the fragmentation measure before analyzing data, a takeaway from the discussion of diversity measures is that the fragmentation of local information can be operationalized in the way that beta diversity is quantified because the conceptual dimensions of both constructs are consistent.[22]

4.5 Descriptive Analysis

The beta diversity measures are calculated based on the number of events in each source and their duplications between them for each city and month. Figure 4.8 shows the pair-wise correlations between beta diversity measures for the 3-month data of 28 urban areas. In the correlation table, β_{GL} is opposite to the direction of β_I, β_E, and β_R, while their correlations are very high. It is because the event overlap rates between sources are generally very low in the dataset. β_{GL}, an extreme measure of volume balance, tends to be higher when the volumes are skewed (i.e., unbalanced) while β_I, β_E and β_R are higher when volumes are more balanced (see Figure 4.7).

The high correlations between these four beta measures ($R^2 \geq 0.94$, $p < 0.001$) show that, even if they put weights differently on balance and overlap, the degree of fragmentation does not vary that much depending on the change in overlaps. In other words, the variations of overlaps in the local event datasets are too small to

[22] While the term *diversity* implies positive meaning, the term *fragmentation* provides a negative nuance. However, when discussing the names of the measures, this study does not consider the nuance differences of the terms but follows academic terms that are used in previous work. After sensitivity analysis, the final measure is named "fragmentation" for this study, because this term signifies potentially-negative effects on people's information inequality.

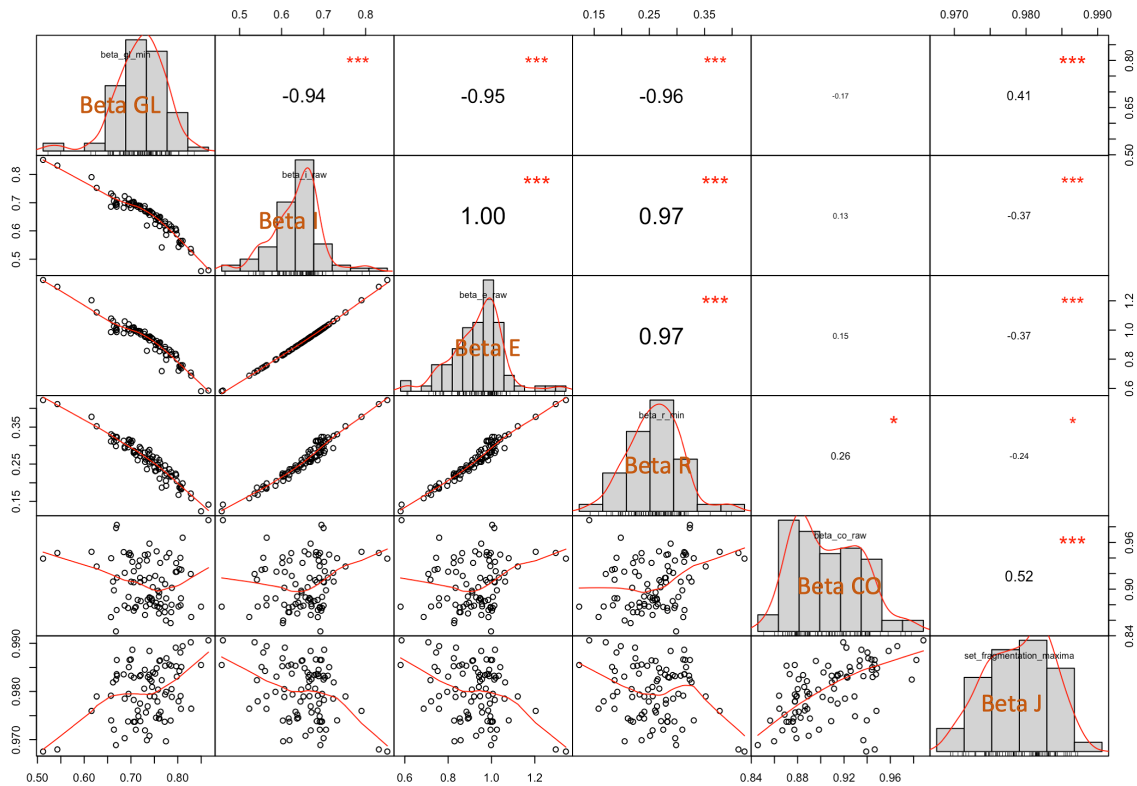

Figure 4.8: Correlations between different beta diversity measures for the 3-month data of 28 urban areas.

be reflected in these four measures. This indicates that these four measures might not properly measure the fragmentation of local information, because they cannot quantify the changes of event overlaps within an urban area over time and between different urban areas.

Meanwhile, β_{CO} is not highly correlated with either the four beta diversity measures (p is N.S.) nor β_J ($R^2 = 0.52$). Because β_J measures the degree of overlap only, it is obvious that β_J is conceptually not a good measure for the fragmentation of local information. This leaves us β_{CO} as potentially the best measure for the fragmentation of local information, because conceptually it takes both overlap and balance into account; methodologically it varies enough depending on both overlap and balance with regard to the event datasets.

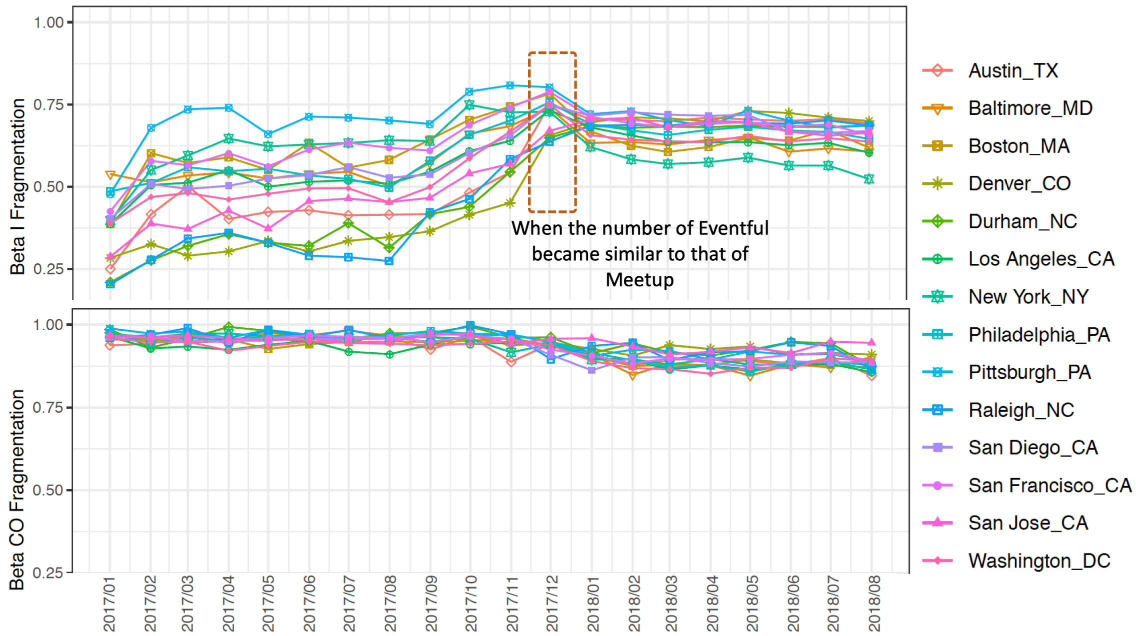

Figure 4.9: Month-to-month changes of β_I and β_{CO} in 14 urban areas over 20 months. Both measures range from 0 to 1 where 0 means the complete consolidation of data and 1 means complete fragmentation.

To see the differences between β_{CO} and other measures, the changes of the fragmentation measures are plotted in Figure 4.9. The graphs show that β_I, which puts more weight on balance for the given dataset, peaks in December 2017. Eventful has significantly increased the amount of data that they curated over time. December 2017 is the time when the number of Meetup events and that of Eventful events becomes similar and the number of Eventful events becomes larger than Meetup's after that (this pattern can be observed in Figure 4.5 and is consistent across different urban areas). Because the balances of event volumes vary significantly due to the change in the amount of Eventful's data, β_I's also vary significantly between 0.2 and 0.8. This variation in the fragmentation of local information marginally reflects the change in event overlaps and, eventually, results in the measurement bias. Intuitively, a large amount of data consolidation is more likely to decrease the

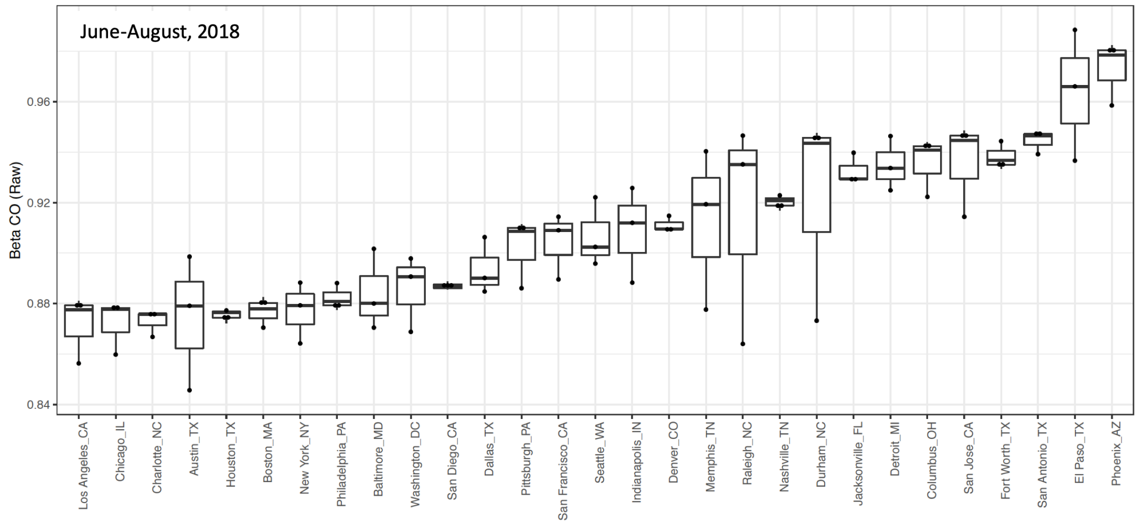

Figure 4.10: The fragmentations of local information (β_{CO}) over 3 months in 28 urban areas.

fragmentation of local information; however, β_I increased in 2018 compared to 2017.

β_{CO} shows more stable changes of the fragmentation scores over time. This measure does not fluctuate too much depending on the change of event volume balance; rather, it puts more weight on overlap compared to β_I. Although the overall variations of β_{CO} are not big, the patterns present better face validity in the fact that the fragmentation rates decrease in 2018 compared to 2017.

Having β_{CO} as the fragmentation measure, the fragmentations of local information in 28 urban areas over 3 months are presented in Figure 4.10 as a box plot. This graph shows between-city and within-city variations in the fragmentation of local information. Urban areas such as Los Angeles, Chicago, Charlotte, Austin, and Houston show relatively low fragmentation rates in their local information landscapes. Conversely, urban areas such as Phoenix, El Paso, San Antonio, and Forth Worth show high fragmentation rates in local information landscapes. Also, urban areas such as Raleigh, Durham, and Austin show high variations over the three months while other urban areas are relatively stable. These characteristics might be related to people's trust, social cohesion, and interactions between an urban area's sub-communities such as ethnic, cultural, and age groups. Therefore, the relationships between the fragmentation of local information and other community characteristics are examined in the next chapter.

4.6 Discussion and Limitations

While the measurement selection and development process are systematic, there are some limitations in the variables generated from this process. The selection of information sources could present a sampling bias. Ideally, calculating the fragmentation measures based on preferably many, randomly-selected, information sources is ideal from an experimental design perspective. However, because it is practically difficult to collect data from many different information sources that exist consistently throughout different cities, the fragmentation scores are generated based on only three sources. If there are many other information sources that are used more frequently by residents than these three sources, the generated variables would present biases in operationalizing the community-level features. It is possible to adjust the variables using post-stratification methods, but it could still be limited because the volumes and dynamics of the rest of the information sources need to be somehow speculated in the stratification process. Due to this reason, it is desirable to collect or predict data from other information sources for future studies on this topic.

Also, it is important to note that β_{CO} is *not* a stable, time-invariant measure for the fragmentation of local event information. For the collected datasets that this study use, β_{CO} is the best because it puts appropriate weights on both balance and overlap. However, it is possible that future information landscapes change dramatically and the variations of the event overlaps could become vary large, both temporally and spatially. In this case, β_{CO} cannot be used continuously because it could put too much weight on overlap over balance. Because of this, future studies need to keep descriptively examining the balance and overlap of local information to better capture the fragmentation of local information. Furthermore, it will be the best if a universal fragmentation formula that includes parameters is developed, so that the weights between balance and overlap can be more systematically considered using the formula.

Despite these limitations, this chapter provides a systematic framework to determine the measure for the fragmentation of local information by adapting the conceptualizations and metrics of diversity that were mainly developed in the field of ecology. While the process and metrics do not provide fine-resolution control over the weights on balance and overlap, this study can be used as a baseline approach to further develop the measurements for the fragmentation of local information.

Chapter 5: Analysis and Results

Using the measures for the fragmentation of local information, a key variable for this study, this chapter conducts statistical analyses to test the hypotheses presented in Chapter 3. Because there are other key variables such as socio-economic deprivation and cultural activity diversity, this chapter first presents the variable generation processes. These processes involve computational treatments and modelings. Then, analytical models, which take the degree of freedom and effect size into account, are constructed to test the hypotheses. Based on the results of the regression analyses, this chapter discusses the implications of the findings.

5.1 Measuring Cultural Activity Diversity

Cultural constructs have been operationalized in many different ways in computational and social sciences as reviewed in Chapter 3, depending on the characterization of a concept, types of data used, and analytical model. For example, Hristova et al. [2018] quantified "cultural capital diversity" using Shannon entropy based on the standardized richness of cultural tags in each neighborhood (i.e., cultural capital). Cultural tags were extracted from Flickr image tags and filtered based on taxonomies generated from Wikipedia Article Tree. From a conceptual perspective, this approach is based on a resource-based view on the diversity of people's perceived urban scenes in a neighborhood.[1] In a similar way but different in its cultural dimension, a study in GIS quantified cultural diversity based on ethnoracial composition in a neighborhood using Shannon entropy [Hu et al., 2019], following the tradition of sociology and political science [Costa and Kahn, 2003, Putnam, 2007, Sturgis et al., 2014].[2]

Unlike this approach, some studies on musical cultures computationally modeled people's cultural tastes at the individual level [Park et al., 2015] and country level [Park et al., 2017]. The former operationalized the diversity of an individual's tastes for music genres (or the degree of boundary spanning) using Rao-Stirling index [Stirling, 2007], which considers not only the balance and variety of one's music consumptions across existing music genres, but also the distances between genres in the conceptual space of the YouTube music market. Theoretically, this work can be seen as a study on individual-level *cultural omnivorousness* following the term by Peterson and Kern [1996] and Goldberg et al. [2016].

Meanwhile, the latter quantified a country's cultural openness, the extent to which a country is open to other cultures, based on countries' YouTube consumptions patterns. Particularly, they measured the betweenness centrality and modified closeness centrality as the two dimensions of cultrual openness using the weighted-edge network of countries where common video subscriptions/views between two

---

[1] Image tags as the basis for quantifying cultural assets do not particularly operationalize individual-level cultural omnivorousness or institution-level atypicality, but are more related to visual manifestations of architectural and scenic culture in a city.

[2] Most of the sociology and political science work on ethnic heterogeneity used HHI (or Simpson index) but from a operationalization perspective, they are consistent.

countries were used to calculate the weight of edges between the two nodes. This study is distinguishable from the former one in the fact that it does not rely on musical genre classifications in quantifying the variety of cultural consumptions; instead, it uses the categories for calculating the cultural similarity between countries. From a theoretical perspective, this study can be seen as ecological work that focuses on the atypicality of a country's culture, the extent to which a country is unique in the ecology of the YouTube video market.

5.1.1 Cultural Omnivorousness and Atypicality in Measurements

While these studies operationalize either cultural omnivorousness based on legitimized genre systems or cultural atypicality in an ecology of actors and artifacts, the concept of "cultural activity diversity" incorporates both genre-specific and ecological characteristics (but does not characterizes people's ethnic or national backgrounds). This concept is reified as the diversity of physical activities that occur in a local community. Conceptually, cultural activities are the physical/social manifestations of both individuals' enactments based on their values/interests within a genre system established in event-driven communities *and* event organizers' strategies on typicality/atypicality for their success in the ecology of local event markets.

In this sense, it is reasonable to use the titles and descriptions of local events to capture this multi-dimensional construct. More importantly, the operationalization of cultural activity diversity needs to take both aspects of the cultural concepts into account. For example, local businesses, a large portion of local event organizers, often need to establish their brands to attract participants (who are potential customers). To do so, they would often organize events that not only span existing genres, but also mix and incorporate diverse cultural elements as part of their strategies to be distinctive. Other small businesses and interest-based event organizers such as photography and concert goers' groups might abide by the established genre systems of a community to benefit from the legitimatized market structure in attracting newcomers [Hannan and Freeman, 1993, Hannan and Carroll, 1992].

From an operationalization perspective, it is challenging to capture the dynamically-changing topics/genres of cultural activities using the pre-defined classifications of events on the websites. In other words, using an established genre system would be useful when focusing only on audience-based diversity. Instead, dynamically-changing genre systems can be better captured through unsupervised approach because this bottom-up approach makes it possible to consider both the audience-side enactments within the established genre systems *and* producer-side strategies to create novel/unique activities that comply less with existing classifications.

Topic modeling algorithms provide such capability in identifying the types of events in an unsupervised way, volumes of activities across the generated topics, and their diversity in a city based on the corpus of event titles and descriptions. Particularly, Latent Dirichlet Allocation (LDA) is used to identify the spectrum of topics available in the event descriptions and titles [Blei et al., 2003]. After event topics are generated, each event is represented by a list of all the topics that have weights. Also, each topic is represented by a list of topical words with weights.

Using these structures of topics, cultural scores are calculated for each topic.

Two topics that have the largest and second-largest weights are selected for each event. Then, the topic with the largest weight adds two points and that with the second-largest weight adds one point. By extracting scores from all the events available in an urban area, this area can be represented with a list of topics where itss elements are the scores that are calculated from the topic distribution of events. For example, the topics of cultural activities in Washington D.C., $C_{Washingon}$ can be represented as a list such as:

$$C_{Washingon} = [t_1 : n_1, t_2 : n_2, ..., t_i : n_i, ..., t_k : n_k]$$

where t_i is the index of the ith topic, k is the number of topics that is used in LDA, and n_i is the aggregate score of topic t_i.

5.1.2 Determining the Number of Topics

To run the LDA against the corpus of event titles and descriptions, it is necessary to decide the number of topics, a key parameter of the LDA algorithm. Because this study does not rely on the genre systems that are defined by the event-based websites, there need to be a systematic way to determine the number of topics. Specifically, internal and external validity tests are required to ensure the number and spectrum of topics.

Internal validity is tested using LDA benchmarking algorithms developed by Arun et al. [2010], Cao et al. [2009], Deveaud et al. [2014], and Griffiths and Steyvers [2004]. These algorithms are implemented in the *ldatuning* package in R [Nikita, 2019]. Griffiths and Steyvers [2004]'s algorithm finds the best number of topics by estimating the posterior probability of the corpus $\mathbf{w}$ (or the likelihood of the data) given the number of topics N_T, thus, finding N_T that maximizes $P(\mathbf{w}|N_T)$ (after fixing priors). Arun et al. [2010]'s algorithm assesses the optimality of the number of topics by factorizing generated topics into two stochastic matrices (i.e., a matrix of words distribution for each topic and that of topics distribution for each document). Then, this algorithm examines how much these matrices diverge using the symmetric KL divergence, given a number of topics. Cao et al. [2009] uses the structure and density of topics by measuring the distances between topics, which is a similar approach for determining K in the K-means clustering. Similarly, Deveaud et al. [2014] uses the symmetric KL divergence measure between all pairs of LDA topics to determine the best number of topics.

Because it is practically inefficient to run LDA tuning algorithms against all the event titles and descriptions, random samples of events are used for the LDA model construction. To sample the events for running the algorithms, the number of Yelp records is used as a basis to sample other event datasets. The number of Yelp events for the 29 cities is 61,849. While the entire Yelp events are included in the training set, the same amount of records is randomly sampled from each of the other information sources, which resulted in 180,647 events from the three sources for the tuning/training dataset.[3]

The overall changes of the benchmarks are presented in Figure 5.1 when the

[3]The final number is slightly smaller than $3 * 61,849$ because some cities have smaller numbers of events. In this case, all the event records are used from the city.

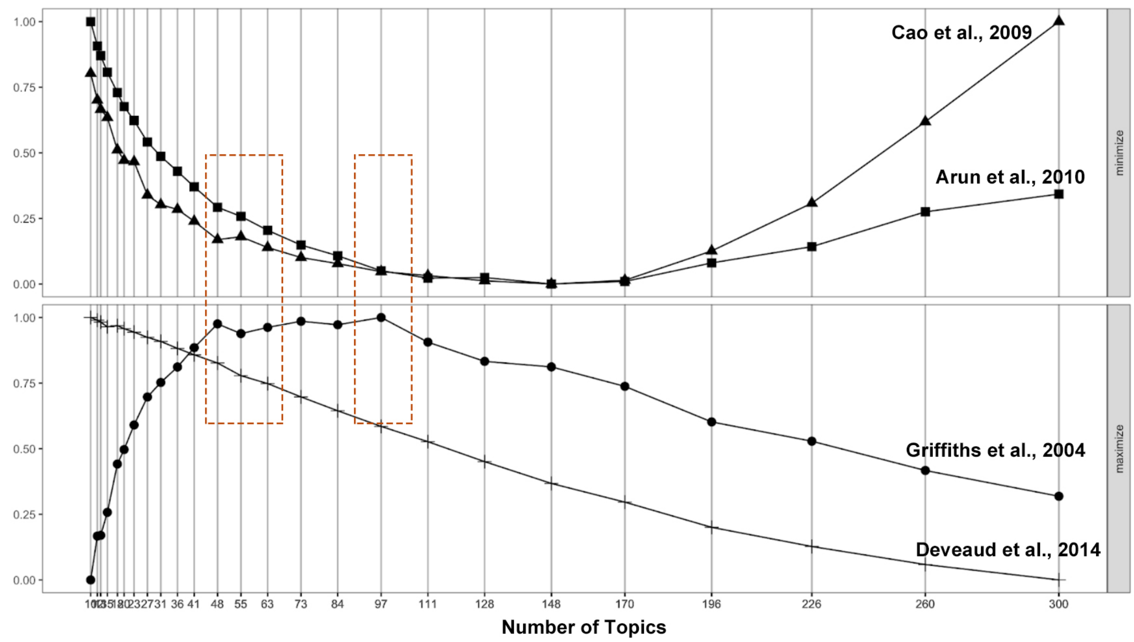

Figure 5.1: Changes of LDA benchmarks when the number of topics changes in a log scale. The dashed-line rectangles indicate local optimum numbers from an internal validity perspective.

number of topics changes in a log scale from 10 to 300. The dashed-line rectangles in the figure indicates the local optimum numbers in terms of internal validity. Because the locally-optimal numbers of topics are identified in the log-scale benchmark plot, these two ranges of optimal numbers are tested to find the number of topics in a granular resolution (i.e., $N_T = 40 - 60 and 80 - 110$) and plotted in Figure 5.2 and Figure 5.3, respectively. The benchmark plots show that potentially optimal numbers of topics are 48, 60, 90, and 96 from an internal validity standpoint.

Using these potential numbers of topics, perplexity scores, a representative LDA performance score given the Dirichlet parameters, α and β, are plotted to see what the best priors are [Manning et al., 1999].[4] Because 90 and 96 are very close to

[4] α is a parameter (or *Dirichlet prior*) that characterizes the topics distribution over documents and β is a parameter that characterizes the distribution of words over topics. For more information about the priors in LDA algorithms, see [Blei et al., 2003].

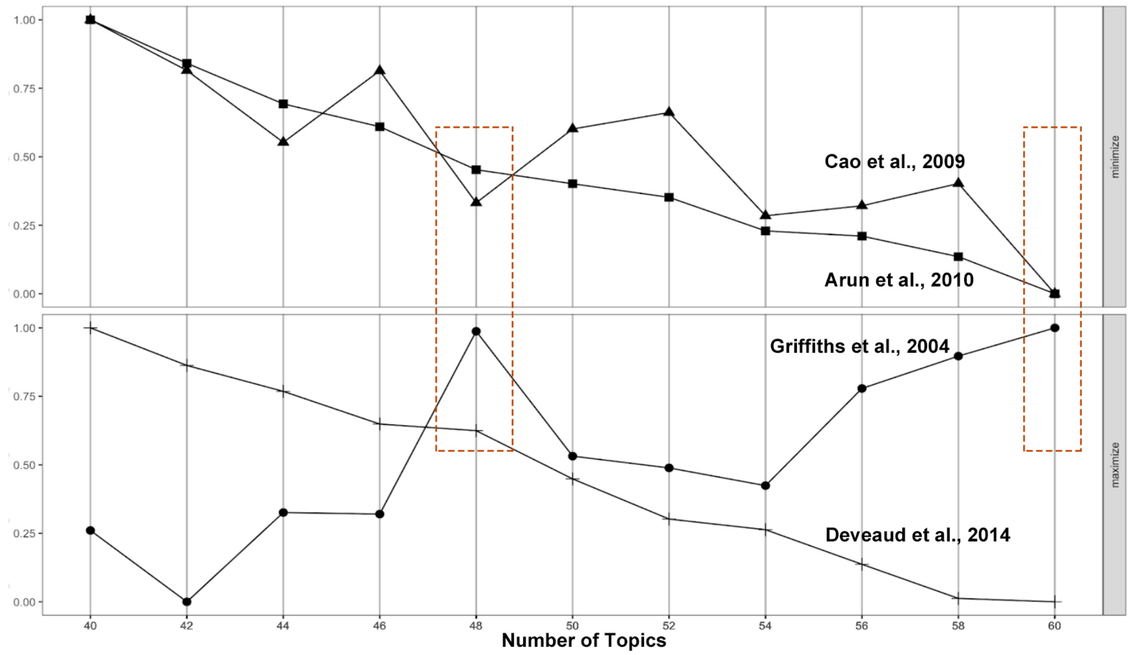

Figure 5.2: Changes of LDA benchmarks when the number of topics changes between 40 and 60. The dashed-line rectangles indicate local optimum numbers from an internal validity perspective.

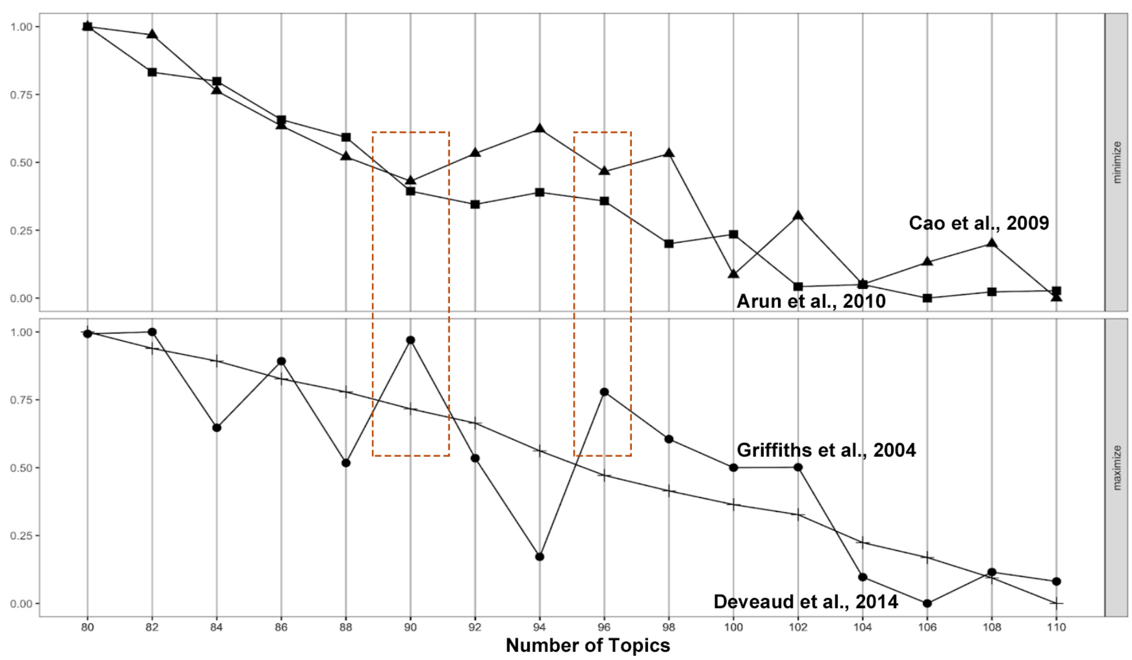

Figure 5.3: Changes of LDA benchmarks when the number of topics changes between 80 and 110. The dashed-line rectangles indicate local optimum numbers from an internal validity perspective.

each other, 96 is omitted from the perplexity tests. From each test, the best pair of priors (i.e., α and β) is selected and topics are generated based on these parameters.

Then, the results of LDA are qualitatively examined to ensure the external validity. For testing external validity, the LDA results are qualitatively examined to see whether cultural topics are grouped and classified reasonably across different topics. Heuristics for the qualitative examination are that the outcome of the LDA algorithm needs to (1) provide a good spectrum of topics that cover the variety of events, (2) present reasonable groupings of words that are exclusive to each other, and (3) distinguish subcategories of general topics so that participants' socio-economic characteristics can vary depending on topics. Especially (3) is both methodologically and theoretically important because this study tests whether socio-economic deprivation affects cultural activity diversity. The classifications of events need to be structured in the way that can show some variations within the participants in a city.

For the qualitative examination of the LDA results, the visual analytics tool, *LDAVis*, developed by Sievert and Shirley [2014], is used because this tool makes it easy to quickly examine the word distribution in each topic by allowing the adjustment of the words relevance score for each topic (e.g., how much a word shows up more in a particular topic compared to other topics). Through this process, the number of topics is determined as 48 ($\alpha = 0.2$ and $\beta = 0.07$). Perplexity plots and a screen shot of the visual analytics tool are presented in Appendix B.

5.1.3 Topic Modeling

After the number of topics and priors are determined based on the 180,647 samples of events, a topic model is trained using the descriptions and titles of these samples (after preprocessing, 179,858 records in total are used for training the model). Then, this model is used to predict other events' topics. In other words, the trained LDA model is used to predict how each event's description and title are represented as a list of topics and their distribution. Specifically, the title and description of each event are concatenated after preprocessing them (e.g., removing stop words and special characters, stemming words, etc.). These aggregated words are tokenized and vectorized as a document term matrix (DTM) in *R*. Finally, the topic modeling algorithm is run using the *text2vec* package that makes use of the WarpLDA algorithm, an $O(1)$-complexity sampling algorithm to run LDA [Chen et al., 2016, Selivanov and Wang, 2018].[5] The results of topic modeling and the predictions based this model are used as the basis for calculating cultural activity diversity for each city. The list of the generated topics and their words distributions are presented in Appendix C.

5.1.4 The Measurement of Cultural Activity Diversity

Although the list-based operationalization makes it possible to understand the cultural activity diversity of a city in a computational way, the similarity and difference between topics could bias the computational model. For example, let us assume

[5] Due to the massive amount of event descriptions, WarpLDA is used instead of brute force computations.

three different topics: jazz concerts, live guitars in the bar, and political movements. In a conceptual space of cultural activities, jazz concerts are more similar to live guitars in the bar than to political movements. In this case, using well-known diversity measures such as Shannon entropy or HHI for measuring cultural activity diversity could present a measurement bias. Using these measures treats jazz concerts and live guitars as being orthogonal to each other, just like the relationship between jazz concerts and political movements. Because of this potential measurement bias, scholars considered the distances between categories in the measurement in addition to balance and variety of entities across topics [Park et al., 2015, Goldberg et al., 2016, Pontikes and Hannan, 2014].

In a similar way, this study also takes the distances between topics into account. Particularly, the distance between two topics is measured based on cosine similarity. Because each topic is represented as a list of topical words and their distribution, it is possible to calculate the distance between two topics, $d_{i,j}$ (where i and j are the indices of two topics, T_i and T_j, respectively), with the angular form of cosine similarity, which is formulated as follows.[6]

$$d_{i,j} = 2 \cdot \frac{cos^{-1}(\frac{T_i \cdot T_j}{|T_i||T_j|})}{\pi} \tag{5.1}$$

Based on the topic modeling results that include the volumes of events across different topics and distances between topics based on cosine similarity, Rao-Stirling diversity index, D, is used as the cultural activity diversity measure, because it not

[6] $1 - \frac{T_i \cdot T_j}{|T_i||T_j|}$ is not used for measuring distance because this measure does not have the triangle inequality property.

only takes variety and balance into account (which is similar to Shannon entropy or HHI) but also considers disparity between topics in the measure [Rao, 1982, Stirling, 2007, Park et al., 2015].

$$D = \sum_{i,j(i \neq j)} (d_{i,j})^{\alpha} \cdot (p_i \cdot p_j)^{\beta} \tag{5.2}$$

where p_i and p_j are proportional representations of topic T_i and T_j, respectively, of an urban area (i.e., balance of cultural activity volumes between two topics), and $d_{i,j}$ is the degree of difference attributed to these two topics (i.e., disparity between two topics). α and β are weighting coefficients that can be used to weight differently on balance and disparity. By default, α and β are both 1. Cultural activity diversity is calculated for each urban area and each month to show its variations between urban areas over time. Cultural activity diversity scores for the target areas over time are presented in Appendix D.

5.2 Variables

5.2.1 Key Variables

Key variables generated or collected for operationalizing the constructs in Figure 3.1 are presented in Table 5.1. The ethnoracial heterogeneity of an urban area is measured using the Herfindahl-Hirschman Index (HHI) following the measures used in previous work [Putnam, 2007, Sturgis et al., 2014]. HHI operationalizes the degree of entity concentrations across a set of categories [Hirschman, 1964]. The

HHI score that is used in previous work on community heterogeneity is a reversed form of HHI, which is also known as Gini-Simpson index.

$$H = 1 - \sum_{i=1}^{N} p_i^2 \tag{5.3}$$

where N is the number of ethnoracial groups and p_i is the share of ethnoracial group i. Socio-economic deprivation is measured using five indicators of deprivation. "Poverty level," one of the five deprivation indicators, in Table 5.1 is the thresholds that are determined by U.S. Census Bureau based on family characteristics and is subject to change every year.[7] To generate deprivation index, Principal Component Analysis (PCA) is used against these five indicators. The first component of the results (PC1) is the deprivation index, which represents 60.45% of the five indicators.

Community engagement is measured using three different variables as shown in Figure 5.1. Because standardized numbers of events and RSVPs are community engagement indicators that are tailored to the online community users, it could be biased in terms of user characteristics. Especially, RSVP information is collected from only Meetup and Yelp, but not from Eventful, because Eventful usually provides links to official RSVP pages such as EventBrite, Meetup, and Yelp rather than being used as a primary RSVPing website.

Also, this study assumes that people often do not RSVP for a same event on multiple websites. This assumption and the linking mechanisms of Eventful suggest that the number of RSVPs from the two sources does not need to be stratified to

[7] See here for more detail: `https://www.census.gov/topics/income-poverty/poverty/guidance/poverty-measures.html`

Construct	Variable(s)	Measurement
Ethnoracial heterogeneity	Ethnoracial heterogeneity	HHI (Gini-Simpson index) of Six ethnoracial groups
Socio-economic segregation	Income inequality	Gini index
Socio-economic deprivation	Percentage of population below poverty level	$\frac{N_{below\_poverty}}{N_{population}}$
	Percentage of rented households (compared to owned)	$\frac{N_{rented\_households}}{N_{households}}$
	Unemployment rate below poverty level	$\frac{N_{unemploy\_below\_poverty}}{N_{population}}$
	Unemployment rate at and above poverty level	$\frac{N_{unemploy\_above\_poverty}}{N_{population}}$
	Percent of people who holds less than bachelor's degree	$\frac{N_{less\_than\_bachelor}}{N_{population}}$
	Deprivation Index	First component (PC1) of PCA against five variables above
Socio-demo. features	Rate of adults over 65	$N_{ppl\_over\_65}/N_{population}$
	Median age	AGE_{median}
	Gender balance	$N_{female}/N_{population}$
The fragmentation of local information	The fragmentation of event data across different websites	β_{CO}
Cultural activity diversity	Local events diversity	Rao-Stirling index
		α-weighted Rao-Stirling
		β-weighted Rao-Stirling
		Shannon index
		HHI
Community engagement	Standardized number of events	$N_{total\_events}/N_{population}$
	Standardized RSVPs per event	$N_{total\_RSVPs}/N_{events}$
	Standardized RSVPs per person	$N_{total\_RSVPs}/N_{population}$

Table 5.1: Key variables that represent constructs in the study and their measurements.

follow the population distribution. In other words, RSVPs in the dataset are not random samples of RSVPs from the entire information sources available in an urban area; rather, they operationalize people's intention to participate in offline events affected by online information.

Cultural activity diversity is measured using five different diversity measures. Although using Rao-Stirling index is conceptually reasonable, it is unclear how the disparity of topics and balance of event volumes across topics should be given. By using five different measures that are slightly different in their components and weighting, it is possible to examine the sensitivity of the cultural activity diversity measures through the analysis.

5.2.2 Control Variables

Control variables (CV) follow the variables used in previous work on community heterogeneity and engagement. Because many of the previous work used individual-level survey data, control variables that are tailored to individuals such as whether one owns a house are not used. All other community-level control variables are used. In addition, some community-level variables that could give rise to the dependent variables are included. For example, the total number of population might matter for cultural activity diversity, because local events are more likely to be diverse in a city with a large number of people than that with a small number of population. In this case, the population needs to be controlled to correctly see the effects of independent variables on cultural activity diversity. In such reason, some

Construct	Variable(s)	Measurement
City population	The number of population in a city	$N_{population}$
# Events	The number of events in a city	N_{events}
Commute time	Average commuting time of people in a city	Average commute time (min)
Ethnoracial group rate[a]	White percentage	$N_{white}/N_{population}$
	Black/African-American percentage	$N_{black}/N_{population}$
	Native American or Alaskan percentage	$N_{native}/N_{population}$
	Asian percentage	$N_{asian}/N_{population}$
	Hawaiian/Pacific Islanders percentage	$N_{pacific}/N_{population}$
	Other ethnic groups percentage	$N_{other}/N_{population}$
Foreign-born Population	Foreign-born population rate	$N_{foreign}/N_{population}$
Citizen rate	U.S. citizen population rate	$N_{citizen}/N_{population}$

Table 5.2: Control variables used in the study. The variables are selectively used depending on the analytical model.

[a] Ethnic group classifications follow those of U.S. Census Bureau.

population-level variables are included when needed. Overall CVs are presented in Table 5.2.

5.3 Descriptive Statistics

Descriptive statistics for the key and control variables are presented in Appendix E. The means and standard deviations of census data-based variables are available in Table E.1. Statistics of event data-based variables for 14 urban areas and 29 urban areas are presented in Table E.2 and E.3, respectively.

Deprivation index generated from the PCA of the five deprivation variables represents about 60% of the five variables. The correlations between these variables are shown in Figure 5.4. This correlation table shows that generated deprivation index represents the socio-economic deprivation well. Population deprivation index is the variable generated from the PCA against the variables of the entire urban areas; meanwhile, deprivation index used in the analysis is generated from the PCA against the datasets for target urban areas.

Deprivation index is positively correlated to Gini index, a socio-economic inequality measure, and negatively correlated to ethnoracial heterogeneity. While the relationship between socio-economic deprivation and ethnoracial heterogeneity is one of the research questions that need to be tested by including other control variables, the simple Pearson correlation between these two variables provides a baseline understanding about their characteristics.

While there is no temporal resolution for the 2017 ACS data used in this study (because the variables are 1-year estimates), variables generated from local event datasets have a relatively-high temporal resolution (i.e., 1-month). Because of this temporal resolution difference, it is necessary to check if there are any systematic changes in variables only because of time. Particularly, the key variables such as the fragmentation of local information, local events diversity, and standardized event participation rates are examined. If there is any systematic change over time in variables, the time variable needs to be considered in the analytical models. Effect size, Hedeg's G is used to show month-to-month changes of each variable and plotted in Appendix F.

There are some longitudinal changes in part of the cultural activity diversity

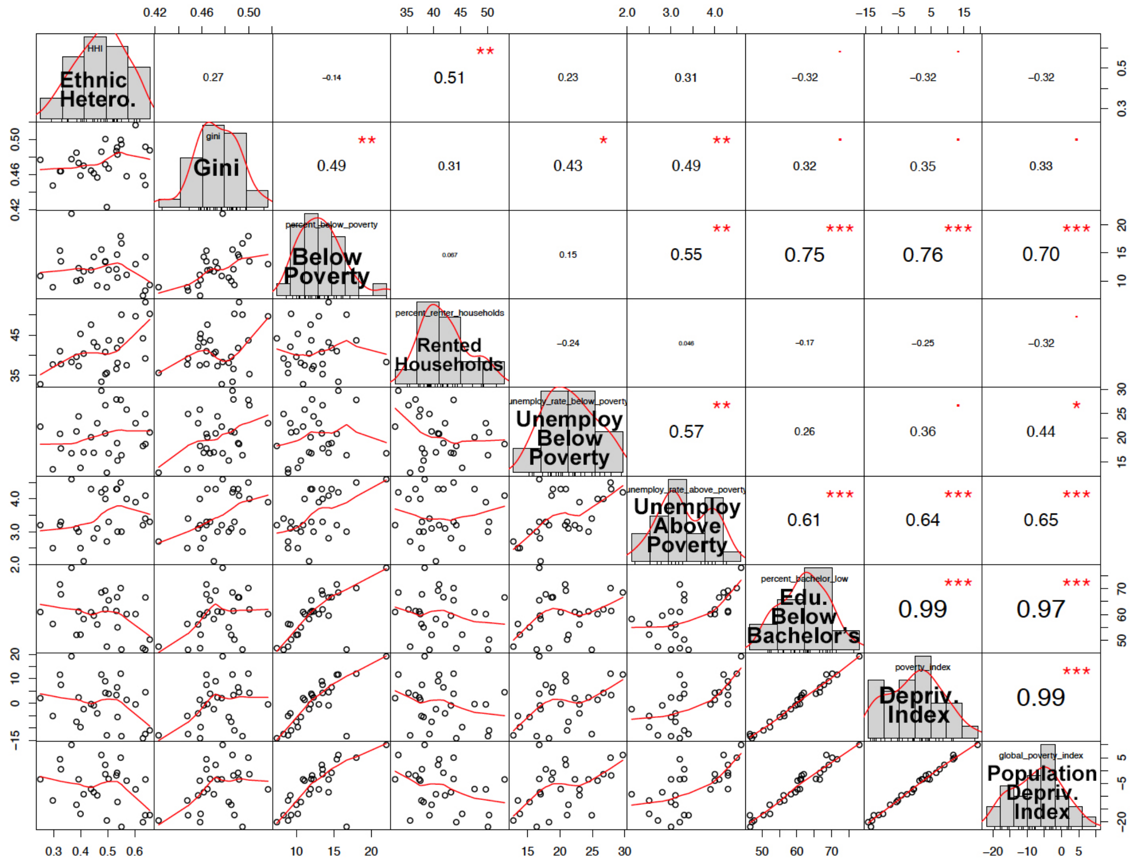

Figure 5.4: Correlations of socio-economic deprivation variables.

and community engagement variables as can be seen in a few graphs of effect sizes (e.g., Figure F.3 or Figure F.5). However, these are not a linear change over time. They are seasonal or irregular changes. Other variables such as the fragmentation of local information do not show meaningful changes from a temporal perspective. This suggests that the time variable is not considered as a factor that linearly affects community characteristics; rather, it can be used as an indirect indicator of event organizers' and website owners' strategies and behaviors across different cities. Especially, the fact that Eventful significantly increased the amount of curated data over time indicates that there have been some systematic changes in the event information landscapes across different urban areas, which needs to be somehow taken into account in the analytical models as a categorical variable.

5.4 Analytical Models and Results

A challenge in the data analysis comes from the number of data points. Because the unit of analysis for all the models is an urban area, the numbers of data points in regressions are too small to include many control variables. Increasing the number of control variables often leads to the overfitting issue. To deal with this challenge, this study uses two different approaches. The first one is the use of Bayesian-based regressions to run linear multi-level models (LMM) in most hypothesis tests (except for H1-1, which uses a large number of data points to provide a baseline). Unlike regular LMMs, this approach minimizes the overfitting or multicollinearity issues by setting priors in estimating the distributions of parameters.

The second strategy is to selectively remove low-effect control variables from the model. Although Bayesian-based LMM is strong in dealing with overfitting, it is not without limit in the number of independent variables. To select the variable(s) that can be removed from the model when necessary, baseline regressions are conducted against all the control variables (from Table G.3 through G.8). Based on these baseline tables, one or two control variables that have least effects on the DV are removed, only if the number of variables goes beyond the capacity of the Bayesian-based LMM. When one or two CVs have to be removed from a model, this change is applied to the other datasets that test the same hypothesis. For example, if a CV of a model testing H1-3 is removed for the 14-city regression, the same CV is removed from the 28-city data as well to ensure the consistency in testing the hypothesis.

Using these models and strategies, the overall summary of regression results is briefly presented in Table 5.3 and 5.4. This table shows the directions of regressions and their significance, based on confidence intervals. If two directional symbols are presented in a cell of the table, it means there are two IVs in the model. If two IVs are present in a model, they are always Gini index and deprivation index, respectively (e.g., "$+-$" means $+$ for Gini index and $-$ for deprivation). Even if some results do not show significant effects in the parameter estimates, the table still shows the directions of parameters. The directions of estimates without 90% confidence interval can still be meaningful in terms of identifying the general patterns of a factor in particular models.[8]

The original datasets for the regressions are the 14-area dataset over 20 months and 28-area dataset over 3 months. Because some differences are found between these two sets of regressions, the same models are tested against the 14-area dataset over 3 months for the same period that the 28-area data covers, so to infer the reasons of the differences. Details on the regressions and their results regarding hypotheses are presented in Appendix G. Specific regression models/methods used in the analyses and the findings regarding the research questions are presented in the following sub-sections.

[8]For details, see Appendix G.

	14 Cities, 3M.		14 Cities, 20M.		28 Cities, 3M.	
	Dir.	CI Sig.	Dir.	CI Sig.	Dir.	CI Sig.
H1-2 SES→Engagement						
Events/population	−−	∗∗	−−	∗∗	−−	∗∗
RSVPs/population	−−	∗∗	−−	∗∗	−−	·∗
RSVPs/event	+−	·∗	−−	·∗	+−	∗∗
H1-3 Ethnic→Engagement						
Events/population	−	∗	−	∗	−	∗
RSVPs/population	−	∗	−	∗	−	∗
RSVPs/event	−	·	−	∗	+	·
H1-5 SES Mediation→Engagement						
Events/population	−	·	−	∗	−	·
RSVPs/population	−	·	−	∗	−	·
RSVPs/event	−	·	−	∗	+	·
H2 SES→Fragmentation						
Fragmentation	−−	·∗	−−	·∗	−+	∗·
Frag. w/ ethnic group ctl.	−−	··	−−	∗∗	−+	∗∗
H3 (1) Fragmentation→Engagement						
Events/population	+	∗	+	·	+	·
RSVPs/population	+	∗	+	∗	+	·
RSVPs/event	−	·	+	·	−	·
H3 (2) Frag.→Engage. w/ SES ctl.						
Events/population	−	·	−	∗	+	·
RSVPs/population	+	·	−	∗	+	∗
RSVPs/event	−	·	+	·	+	·
H4 SES→Cultural Activity Diversity						
Rao-Stirling w/ α or β variations	+−	∗∗	++	∗·	+−	··
Shannon entropy	+−	··	++	∗·	+−	··
HHI for diversity	+−	··	++	∗∗	+−	··
H5 SES*Frag→Cultural Activity Diversity						
Rao-Stirling w/ α or β variations	−−	··	−−	∗·	−−	··
Shannon entropy	−−	··	−−	··	+−	··
HHI for diversity	−−	··	−−	··	+−	··

Table 5.3: An overview of the regression results for all hypotheses except for H1-4. "∗" indicates the Bayesian-based estimation of parameters has more than 90% confidence in estimating posteriors; conversely, "·" means it is less than 90% confidence. If there are two IVs in a model, two symbols in the "Dir." column are presented in the order to that of the "CI Sig." column.

	14 Cities, 3M.		14 Cities, 20M.		28 Cities, 3M.	
	Dir.	CI Sig.	Dir.	CI Sig.	Dir.	CI Sig.
H1-4 Ethnic*Old Gen.→Engagement						
Events/population	−	·	−	∗	−	·
RSVPs/population	−	·	−	∗	+	·
RSVPs/event	−	·	−	∗	+	·
H1-4 Ethnic*Med. Age→Engagement						
Events/population	−	·	−	·	−	·
RSVPs/population	−	·	−	·	+	·
RSVPs/event	−	·	−	·	+	·
H1-4 Ethnic*Gender→Engagement						
Events/population	+	·	+	·	+	·
RSVPs/population	+	·	+	·	+	·
RSVPs/event	+	·	+	·	+	·
H1-4 Gini*Old Gen.→Engagement						
Events/population	+	·	+	·	+	·
RSVPs/population	+	·	−	∗	+	·
RSVPs/event	+	·	−	∗	+	·
H1-4 Gini*Med. Age→Engagement						
Events/population	−	·	−	∗	−	·
RSVPs/population	−	·	−	∗	+	·
RSVPs/event	−	·	−	∗	+	·
H1-4 Gini*Gender→Engagement						
Events/population	+	·	+	∗	+	·
RSVPs/population	+	·	+	∗	+	·
RSVPs/event	+	·	−	·	−	·
H1-4 Dprv*Old Gen.→Engagement						
Events/population	−	·	+	·	−	·
RSVPs/population	−	·	+	∗	−	·
RSVPs/event	−	·	+	∗	−	·
H1-4 Dprv*Med. Age→Engagement						
Events/population	+	·	+	·	−	·
RSVPs/population	−	·	+	∗	−	·
RSVPs/event	−	·	+	∗	−	·
H1-4 Dprv*Gender→Engagement						
Events/population	−	·	−	·	−	·
RSVPs/population	−	·	+	·	−	·
RSVPs/event	−	·	+	·	−	·

Table 5.4: An overview of the regression results for H1-4.

5.4.1 RQ1: A Baseline Analysis on Community Heterogeneity and Socio-economic Deprivation

Four hypotheses under RQ1 (i.e., H1-1 to H1-4) are intended to provide a basis for the subsequent analyses by examining whether the findings from previous work are consistent in the context of this study.

RQ1-1. Hypothesis H1-1 is tested using only the ACS data that includes 269 urban areas in 2017. There are more than 269 urban areas according to the U.S. Census Bureau, but because this dataset is 1-year estimates based on community surveys, the 2017 dataset does not include many small areas in the United States. Also, because the ACS data does not have a temporal resolution within 2017 (i.e., values are per city for the entire year), H1-1 can be tested using multi-level regression models with a nested effect of the state on the dependent variable (i.e., socio-economic features can be nested within the state based on geographical, industrial, and political effects). To examine whether ethnoracial heterogeneity (HHI) predicts deprivation index and Gini index, models are constructed by replicating previous studies (but without individual-level survey measurements) following this form.

$$P_c = \gamma_0 + \gamma_1 H_c + \gamma_2 S_c + + \sum_{i(i>2)} \gamma_i C_{i,c} + \varepsilon_c \tag{5.4}$$

where P_c is deprivation or Gini index of community c; H_c is the HHI of community c, ethnoracial heterogeneity index; $C_{i,c}$ is the ith control variable in community c; and S_c is a categorical variable with random effects for state in which community c

is located.

The first models, Model Gini-1 and Depr-1 (Table G.1), naïvely predict Gini index and deprivation index, respectively, with HHI by controlling only citizenship, age, population, and average commute time. Many researchers raised concerns about predicting civic engagement and social trust using only community heterogeneity due to their conceptual simplicity.[9] As such, Model 5.4 is used to examine the effects of ethnoracial heterogeneity on socio-economic deprivation/inequality, an intermediary variable that previous studies suggested for alleviating the simplicity of the original models about community engagement [Sturgis et al., 2014, Abascal and Baldassarri, 2015b]. The rest of the models includes other variables such as proportions of ethnic groups and deprivation indicators such as unemployment rates. The results of regressions are presented in Table G.1 based on the data of 269 urban areas in the U.S. in 2017.

The regression results of Model Gini-1 and Depr-1 show that ethnoracial heterogeneity is positively correlated to both socio-economic inequality (i.e., Gini index) and deprivation (i.e., deprivation index), thus rejecting the null hypothesis for H1-1. An ethnically-diverse community without considering actual proportions of dominant and minority groups tends to be more economically segregated and poorer. After controlling the proportions of ethnic groups in communities (Model Gini-2 and Depr-2), however, ethnoracial heterogeneity is no longer correlated to both Gini and deprivation indices. This indicates that using ethnoracial heterogeneity

[9] The original form of HHI is intended for operationalizing market concentration in industry. To quantify the heterogeneity or diversity, however, this needs to be reversed as $1 - HHI$, which is also known as Gini-Simpson index. This reversed HHI is used throughout this dissertation.

itself for representing community heterogeneity does not thoroughly operationalize the diversity of ethnic groups in a community; rather, it indirectly represents socio-economic deprivation and inequality (more so for deprivation ($\gamma = .62$) than inequality ($\gamma = .23$)).

To further confirm the relationship between ethnoracial heterogeneity and socio-economic deprivation/inequality, ethnoracial heterogeneity (HHI) is predicted with socio-economic features along with other control variables (Table G.2). The results are consistent with Table G.1 where socio-economic deprivation and inequality are positively correlated to community heterogeneity. A new finding from regressions that use the original deprivation indicators instead of PCA-based deprivation index (Model HHI-2, HHI-4) is that the proportion of the below-poverty-level population in a community, one of the five deprivation indicators, is negatively correlated to community heterogeneity. This suggests that urban areas with many low-income families tend to be geographically segregated from other wealthier areas and ethnically biased toward people of color. This inference is confirmed by (1) the positive correlation between ethnic heterogeneity and minority groups' proportions and (2) the negative correlation between ethnoracial heterogeneity and the proportion of white people in the community.

RQ1-2. Except for H1-1, all the other hypotheses are tested using the 14-area and 28-area datasets because key variables are generated from these datasets. Hypothesis H1-2 is tested to see if socio-economic deprivation/inequality is correlated to community engagement indicators such as people's participation in local events. Because community engagement indicators are based on local event datasets, the temporal granularity of community engagement indicators is higher than ACS-based variables. Because there is no systematic change over time in the community engagement indicators, as shown in Appendix E, the temporal factors are used as an indirect measure for ecological, platform-wide changes in each month. Accordingly, the analytical models are linear multi-level models (LMMs) where event data-based DVs are nested within the time dummy variable.[10]

Unlike the models used in H1-1 for ACS data, state is not used as the random-effect variable because the number of urban areas in each state is too small, limiting the error variations within each factor (see Appendix A for the numbers of cities). Putting all together, community engagement indicator, $E_{c,t}$, is modeled as follows:

$$E_{c,t} = \gamma_0 + \gamma_1 P_c + \gamma_2 I_c + \gamma_3 T_t + \sum_{i(i>3)} \gamma_i C_{i,c} + \varepsilon_{c,t} \tag{5.5}$$

where P_c is the deprivation index of local community c, I_c is the Gini index of local community c, T_t is the random-effect time variable in month t, $C_{i,c}$ is the ith control variable in community c, and ε is the error term. All the community engagement indicators are scaled because some variables are by nature very small to see the actual effect (e.g., RSVP normalized by population is too small due to the large amount of population). For running the regressions, as briefly mentioned, Bayesian-based LMMs are used instead of the regular linear multi-level models because of the small number of data points per each month. Bayesian-based multi-level models estimate fixed- and random-effect parameters, γ's, through Markov Chain Monte

[10]Ch.3,4. [Singer and Willett, 2003]

Carlo (MCMC) rather than using maximum likelihood as regular LMM does [Gabry and Goodrich, 2018].

Following the approach of previous work, Table G.3 and Table G.4 show the baseline statistics on how control variables that represent community characteristics are related to community engagement indicators. These baseline results are used to examine high- and low-impact control variables. Regarding community engagement variables, it is possible to see that the rates of Hawaiians/Pacific islanders and native Americans/Alaskans are less effective across different models and datasets. Based on this observation, these two variables are removed only if the number of variables cannot be handled by the regression model.

Table G.9, G.10, and G.11 present consistent results that both Gini and deprivation indices are negatively correlated to community engagement indicators, particularly for the standardized number of events per population and standardized number of RSVPs per population, across the three datasets. This means that an economically-segregated and deprived community tends to have fewer events organized in the urban area and also fewer participants in existing events. Similarly, the results for the RSVPs per event indicate that people in a highly-deprived community are less likely to show their intentions to participate in local events.

These results are consistent with theoretical findings in the literature. Although previous work on this topic mainly used national surveys for measuring people's trust level toward other groups in the individual level or counted the numbers of local organizations and community board meetings [Abascal and Baldassarri, 2015b, Sturgis et al., 2014, Costa and Kahn, 2003], people's creation of and attendance in local events behave similarly with these conventional indicators of civic engagement and social trust (thus, confirms the external validity of the new DVs). The results of testing H1-2 reaffirms that the theoretical explanations of previous work on the negative relationship between socio-economic inequality/deprivation and community engagement (i.e., the numbers of RSVPs and events normalized by population) are still valid in the context of local events that are organized or acknowledged through online websites.

While regressions against the number of RSVPs per event show a consistent pattern for deprivation index (i.e., negative), its correlation with Gini index is either not significant (for both of the 14-area datasets) or opposite to the others (i.e., positive in the 28-area dataset). Unlike other DVs, the regression results against the 28-area data show that economically-segregated urban areas tend to have a higher number of RSVPs per event for local events, which is contradictory to the hypothesis. Because regressions for both the 20-month and 3-month data in 14 urban areas present non-significant results in predicting the number of RSVPs per event, the positive correlation between Gini index and RSVPs per event might stem from the spatial variations of socio-economic inequality between different urban areas, rather than temporal variations.

Therefore, two different interpretations are possible for the regression results of H-2 against the standardized number of RSVPs. On the one hand, it is possible that there are urban areas that present a high socio-economic inequality but not at the level of an absolute deprivation. In this case, socio-economic inequality might not limit people's access to technology too much (thus, high technology penetration rate); rather, it could diversify the social groups online, which would lead to a higher number of RSVPs per event. This inference is supported by the fact that the positive correlation between RSVPs per event and socio-economic inequality represents the online dynamics rather than the dynamics of local communities themselves.

On the other hand, it is also possible that this counter-intuitive result is a product of seasonal effects. Although not significant, the regression result of the 14-area dataset over 3 months shows the positive correlation as well, just like the results for the 28-area data. This indicates a possibility that local events organized during the summer are less affected by socio-economic inequality, maybe due to the relatively low cost of activities, which shapes the role of economic inequality as a source of diversity.

Whichever the reason for the inconsistent findings is, the negative correlation between Gini index and the number of RSVPs normalized by population shows that people's participation in local events is still low in urban areas with high inequality, indicating that inequality still segregates a high portion of population from attending events outside of the online realm. These findings, overall, suggest that the regressions reject the null hypothesis and support hypothesis H1-2.

R1-3. H1-3 tests whether and how ethnoracial heterogeneity is correlated to community engagement. The regression model used for testing this hypothesis is as follows:

$$E_{c,t} = \gamma_0 + \gamma_1 H_c + \gamma_2 T_t + \sum_{i(i>2)} \gamma_i C_{i,c} + \varepsilon_{c,t} \tag{5.6}$$

where H_c is the ethnogracial heterogeneity (HHI) of local community c, T_t is the random-effect time variable in month t, $C_{i,c}$ is the ith control variable in community c, and ε is the error term.

Table G.12 and G.13 present the regression results. Except for the regressions against the number of RSVPs per event for the 14- and 28-area datasets over 3 months, all other 7 regression results show consistent patterns where these two variables are negatively correlated to each other. This pattern is consistent with the literature that reported the negative effect of community heterogeneity on social cohesion and civic engagement, although some of the community heterogeneity models were oversimplified. In this sense, the regression results reject the null hypothesis and support hypothesis H1-3.

Similar to the R1-2 results, inconsistencies in the results for the number of RSVPs could stem from a couple of reasons. On the one hand, the low confidence level in the regressions against the number of RSVPs per event in the 3-month datasets indicates that the relationship between ethnoracial heterogeneity and community engagement need more data points (potentially due to seasonal effects). On the other hand, it is possible, based on the observation of the opposite directions of estimates between the regressions of the 3-month datasets, that 14 urban areas' dynamics could be shaped by or conflated with technology start-up culture. Because the 14 cities are sampled based on high tech-startup activities, it is possible that people's use of online websites for organizing and attending events is prevalent (i.e., technology penetration in the communities is higher) in the 14 urban areas compared to other metropolitan areas.

While there is explicit theoretical discussions found in the literature about the relationship between ethnoracial heterogeneity and technology entrepreneurship, it is known that tech-savvy populations are often ethnically-biased [Kim-Mai Cutler, 2015]. Controlling for the proportions of ethnic groups in the regression models, due to this reason, might have surfaced the high engagement of people in online-initiated local events in's in the 28 urban areas (based on their diverse cultural backgrounds and tastes). Conversely, due to the control of ethno-demographic variables, the effects of ethno-demographic features are removed from the effects of tech-startup culture that are embedded in the variable of ethnoracial heterogeneity in the 14 urban areas, which resulted in low confidence level in predicting community engagement. This suggests that ethnoracial heterogeneity might play a role as a factor that gives rise to diversity by combining with other factors, the variable itself might have little implications for community engagement. To better understand this dynamic of communities, however, a larger study on the U.S. urban areas with individual-level data will be necessary in the future.

R1-4. Although individual-level participation patterns cannot be detected from the data, H1-4 which examines whether socio-demographic features moderates the relationships (1) between ethnoracial heterogeneity and community engagement (Equation 5.7) and (2) between socio-economic inequality/deprivation and community engagement (Equation 5.8 and 5.9).[11] The analytical models are as follows:

$$E_{c,t} = \gamma_0 + \gamma_1 H_c + \gamma_2 (H_c * D_c) + \gamma_3 D_c + \gamma_4 T_t + \sum_{i(i>4)} \gamma_i C_{i,c} + \varepsilon_{c,t} \tag{5.7}$$

[11] See the summary in Table 5.4 for the regression results. Details of the results are presented in Tables from G.15 through G.41

$$E_{c,t} = \gamma_0 + \gamma_1 P_c + \gamma_2 I_c + \gamma_3 (P_c * D_c) + \gamma_4 D_c + \gamma_5 T_t + \sum_{i(i>5)} \gamma_i C_{i,c} + \varepsilon_{c,t} \qquad (5.8)$$

$$E_{c,t} = \gamma_0 + \gamma_1 P_c + \gamma_2 I_c + \gamma_3 (I_c * D_c) + \gamma_4 D_c + \gamma_5 T_t + \sum_{i(i>5)} \gamma_i C_{i,c} + \varepsilon_{c,t} \qquad (5.9)$$

where H_c is the ethnogracial heterogeneity (HHI) of local community c, P_c is the deprivation index of local community c, I_c is the Gini index of local community c, D_c is the demographic variable for urban area c (i.e., either the rate of adults who are more than 65 years old, median age, or the rate of female population who are more than 16 years old), T_t is the random-effect time variable in month t, $C_{i,c}$ is the ith control variable in community c, and ε is the error term.

A finding from the regression results is that the generational factors, the rate of adults over 65 and median age, negatively moderates the effect of ethnoracial heterogeneity on community engagement in the 14-area data over 20 months, meaning urban areas with a high proportion of older generations tend to worsen the effects of ethnoracial heterogeneity and socio-economic inequality on community engagement (Table G.15, G.16, and G.25). These findings are consistent with previous work that explains the moderating effects of generational factors based on contact/conflict theory regarding social cohesion/trust [Sturgis et al., 2014]. While there is a contact-conflict tension existing in people's perception toward other social/ethnic groups, constant exposure to out-group members increases people's chances to increase social cohesion (theoretically, contact-based forces become dominant in the relationship).

Conversely, a lack of contact with out-group members increases the possibility that people perceive out-group members as threats (thus, conflict-based forces becomes dominant between groups). According to previous work, an urban area with many old-generation adults is more likely to incline toward the latter, which could decrease community engagement at the community level. Although this model and justification are conceptually simplistic, this model is tested as part of examining the confounding effect between ethnoracial heterogeneity and socio-economic inequality/deprivation.

However, all the 3-month datasets both for 14- and 28-area datasets present less than 90% confidence in estimating the posteriors, providing weak evidence for the directions of the correlations. This indicates that the degree of community engagement varies longitudinally rather than spatially with respect to generational factors.

Another finding is that the generational factors positively moderate the effect of socio-economic deprivation on community engagement (Table G.26 and G.17), meaning a community with older generations tends to lessen the negative effect of socio-economic deprivation on community engagement. This implies that, while the generational factors are critical in inter-ethnic-group perceptions about how conflict and contact plays out (as Abascal and Baldassarri [2015b] suggests), generational factors might behave differently in inter-economic-class interactions from a conflict/contact perspective.

Finally, gender balance does not provide statistically significant results except for the positive moderating effect of it between Gini index and community engagement. A higher rate of female over 16 is more likely to lessen the negative effect of socio-economic inequality on the numbers of events and RSVPs that are normalized by population in the 14-area data over 20 months. This indicates that genderoriented factors might play an important role in alleviating inter-class segregation.

Generational and gender-related factors in the context of local events pose important questions for the theories of community engagement. In order to examine the roles of these factors more precisely, however, individual-level studies are necessary in the future.

RQ1-5. Finally, the mediating effect of socio-economic inequality/deprivation between ethnoracial heterogeneity and community engagement is tested (Table G.42, G.43, and G.44). The analytical model for this test is as follows:

$$E_{c,t} = \gamma_0 + \gamma_1 H_c + \gamma_2 P_c + \gamma_3 I_c + \gamma_4 T_t + \sum_{i(i>4)} \gamma_i C_{i,c} + \varepsilon_{c,t} \tag{5.10}$$

where H_c is the ethnogracial heterogeneity (HHI) of local community c, P_c is the deprivation index of local community c, I_c is the Gini index of local community c, T_t is the random-effect time variable in month t, $C_{i,c}$ is the ith control variable in community c, and ε is the error term.

While not significant for the 3-month datasets, the results of the 14-area dataset over 20 months confirm that ethnoracial heterogeneity is significantly correlated to community engagement when socio-economic inequality and deprivation index are controlled. Overall, hypotheses H1-1 through H1-5 test whether ethnoracial heterogeneity mediates or is confounded with socio-economic deprivation/inequality. Conceptually, ethnoracial heterogeneity is confounded with socio-economic deprivation/inequality, which has been discussed in the literature [Sturgis et al., 2014], because the composition of ethnic groups is highly dependent on the history of immigration and politics in U.S. cities.

The series of hypothesis tests for RQ1 confirms that ethnoracial heterogeneity is not meaningful too much by itself without considering socio-economic features in local communities. These findings and implications provide a baseline understanding about the data and a basis for subsequent research questions: the main independent variables are socio-economic deprivation/inequality of a community with respect to the fragmentation of local information landscapes, rather than community heterogeneity *per se*.

5.4.2 RQ2: Socio-Economic Deprivation, Inequality, and The Fragmentation of Local Information

Table G.45 through G.47 show the regression results on whether socio-economic deprivation and inequality are correlated to the fragmentation of local information. Two models are used for testing H2; one with controlling ethnic groups and the other without controlling ethnic groups' proportions. The analytical models for predicting the fragmentation of local information, $F_{c,t}$, follow this form:

$$F_{c,t} = \gamma_0 + \gamma_1 P_c + \gamma_2 I_c + \gamma_3 T_t + \sum_{i(i>3)} \gamma_i C_{i,c} + \varepsilon_{c,t} \tag{5.11}$$

where P_c is the deprivation index of local community c, I_c is the Gini index of local community c, T_t is the random-effect time variable in month t, $C_{i,c}$ is the ith control variable in community c (one of CV sets that (1) contains variables for ethnic groups and (2) that does not contain them), and ε is the error term.

Hypothesis H2 is made based on an inference that assumes that members in highly-segregated communities would have narrower contacts with each other, especially with out-group individuals. In such situations, conflict theory is dominant over contact theory and social trust between social/demographic groups tends to be lower. This would lead to an ad-hoc provision of information to online platforms due to a lack of communication between social/demographic groups, creating a highly-fragmented information landscape.

In all the tests, however, it is found that socio-economic inequality is negatively associated with the fragmentation of local information, which is opposite to hypothesis H2. This indicates that the pure effect of socio-economic inequality after controlling deprivation index indirectly represent (1) the diversity of poeple's cultural tastes and interests (i.e., cultural omnivorousness at the community level) and/or (2) their diversified use of local event platforms. After controlling the ethnic group composition, the negative correlation between socio-economic inequality and the fragmentation of local information becomes stronger, suggesting that the role of socio-economic inequality as diversity or diversified use of technology is even more apparent. This also implies that ethnic group compositions are related to the fragmentation of local information, which partially aligns with the original inference in the hypothesizing process.

When it comes to socio-economic deprivation, the regression results are inconsistent between the 14-area and 28-area datasets. While the regression for the 14-area data shows the negative relationship with the fragmentation of local information, the 28-area data presents the positive correlation between the IV and DV, which supports H2. The result for the 14-area data over 3 months indicates that this inconsistency does not stem from the short period of time (i.e., seasonal effects); rather, the difference might stem from sampling bias. Because the 14 urban areas are sampled based on high tech-startup activities, the cultural norms and social dynamics might be driven by technology-savvy population.

This cultural and industrial effect of a community could increase the technology penetration rate within the communities. If so, people from different socio-economic classes including below-poverty-level residents would have higher access to local information, compared to other cities with low start-up activities. Conversely, the samples of the 28 urban areas are less influenced by tech-startup culture and relatively diverse in their social and cultural boundaries among social/demographic groups. While socio-economic inequality might play a role as a source of people's diverse interests and their diversified technology use, people below the poverty level would be rather more segregated from other groups in urban areas with large populations. This segregation could lead to an ad-hoc creation of information within one's own social group and, if so, would lead to the fragmentation of local information.

Overall, the regression results reject the null hypothesis but support only part of the hypothesis, specifically for socio-economic deprivation for in the 28 urban areas over 3 months. Using a more rigorous, large-scale sampling method and includsion of technology penetration as a moderating variable in the model will clarify the new findings and inferences better in future studies.

5.4.3 RQ3: The Fragmentation of Local Information and Community Engagement

If the pure effect of socio-economic inequality is an indirect proxy for the diversity in people's cultural tastes and interests, the fragmentation of local information might not discourage people from attending local events. This is because the fragmentation of local information could be actually the diversity of information platforms that many people are aware of and use. Although hypothesis H3, which anticipates the negative correlation between the fragmentation of local information and community engagement, is inferred based on the original discussion for hypothesizing H2, the results of socio-economic inequality in the H2 tests imply a different meaning and impact of the fragmentation of local information.

The analytical models used to test H3, one without controlling socio-economic factors and the other with controlling them, are as follows:

$$E_{c,t} = \gamma_0 + \gamma_1 F_{c,t} + \gamma_2 T_t + \sum_{i(i>2)} \gamma_i C_{i,c} + \varepsilon_{c,t} \tag{5.12}$$

$$E_{c,t} = \gamma_0 + \gamma_1 F_{c,t} + \gamma_2 P_c + \gamma_3 I_c + \gamma_4 T_t + \sum_{i(i>4)} \gamma_i C_{i,c} + \varepsilon_{c,t} \tag{5.13}$$

where $F_{c,t}$ is the fragmentation of local information (β_{CO}) in community c in month t, P_c is the deprivation index of local community c, I_c is the Gini index of local community c, T_t is the random-effect time variable in month t, $C_{i,c}$ is the ith control variable in community c, and ε is the error term.

Table G.48, G.50, and G.52 present the regression results. While some tests do not provide statistically-significant results, those that are significant (i.e., 14 urban areas' numbers of events and RSVPs normalized by population) show the positive correlations between the fragmentation of local information and community engagement. While opposite to the original inference of H3, this result and its implication are consistent with those of H2. If the fragmentation of local information is a result of an increased diversity in a community and their diversified use of technology which stem from a high technology penetration rate (as H2 results imply), the diverse uses and consumptions of information platforms would give rise to people's participation and creation of local events.

Although these findings provide a new insight on the meaning of the fragmentation of local information, the initial tests do not consider socio-economic features of the communities. Because socio-economic deprivation is sometimes positively correlated to the fragmentation of local information, controlling these variables could change the results. As such, Table G.49, G.51, and G.53 present the regression results after controlling the socio-economic deprivation/inequality in the target urban areas. Controlling these variables changes the directions of the estimates in the 14-area dataset, suggesting that the pure effect of the fragmentation could be detrimental to community engagement. However, the inclusion of socio-economic features does not change the results for the 28-area dataset.

This change in the directions after controlling socio-economic features is consistent with the discussion and implications in RQ2. The 14 urban areas have a higher technology penetration rate across different socio-economic groups. Because the effects of deprived groups' technology use were already embedded in the effect size of information fragmentation in the original model, controlling deprivation would surface the pure effect of information fragmentation, which is detrimental to community engagement. Furthermore, controlling socio-economic inequality could eliminate the effect of diversity that stems from it, thus give rise to the effect size of the fragmentation of local information. Meanwhile, in 28 urban areas, socio-economic inequality and deprivation have conflicting effects on the fragmentation of local information, because inequality does not necessarily mean the diversified use of technology. These two effects in the model would conflict with each other and offset the overall effects, resulting in minimal changes in the estimates.

Overall, the results partially reject the null hypothesis, supporting part of the hypothesis only when socio-economic inequality/deprivation is considered. When not considering socio-economic features, the meaning of the fragmentation of local information can change because it could signify the extent to which people use diverse technological platforms for acquiring and creating local event information.

5.4.4 RQ4: Socio-Economic Deprivation and Cultural Activity Diversity

H4 tests whether and how socio-economic features are related to cultural activity diversity. The analytical models for predicting cultural activity diversity, $R_{c,t}$, follow this form:

$$R_{c,t} = \gamma_0 + \gamma_1 P_c + \gamma_2 I_c + \gamma_3 T_t + \sum_{i(i>3)} \gamma_i C_{i,c} + \varepsilon_{c,t} \tag{5.14}$$

where P_c is the deprivation index of local community c, I_c is the Gini index of local community c, T_t is the random-effect time variable in month t, $C_{i,c}$ is the ith control variable in community c, and ε is the error term.

Table G.54 shows a positive correlation between socio-economic inequality and cultural activity diversity for the 14-area datasets over both 3 and 20 months. This result is consistent with the discussions in RQ2 and RQ3 where the pure effect of socio-economic inequality can be a source of people's diverse tastes and interests in cultural activities and their diversified use of technology. Because the 14 urban areas are sampled based on high tech-startup activities, the cultural norms and people's use of technology in these communities could be shaped by technology start-ups, as the results of H2 and H3 tests suggest.

This implication is supported by the regression result against socio-economic deprivation for the 14 areas over 20 months. Because of the high technology penetration rate in the 14 urban areas, socially- and economically-deprived groups might have better access to technological platforms. In this case, socio-economic deprivation might also play a role as the diversified use of technical platforms based on diverse socio-economic classes' increased access to technology, just like socio-economic inequality.

Meanwhile, the regression results for the 3-month datasets from both 14- and

28 urban areas present the negative correlations between socio-economic deprivation and cultural activity diversity, even though some results are not above the 90% confidence interval (Table G.55 and G.56). On the one hand, these patterns indicate that the diversity of cultural activities varies between urban areas (i.e., spatially) while negatively affected by socio-economic inequality/deprivation. On the other hand, cultural activity diversity is positively correlated to socio-economic inequality/deprivation from a temporal perspective. In other words, there might be a seasonal effect in cultural activity diversity with regards to socio-economic features.

While the regression results reject the null hypothesis, the directions of the main effects are opposite to H4, not supporting the original hypothesis but supporting the new explanations discussed in RQ2 and RQ3.

5.4.5 RQ5: Moderating Effects of The Fragmentation of Local Information

H5 tests whether and how the fragmentation of local information moderates the effect of socio-economic deprivation/inequality on cultural activity diversity. The analytical models are as follows:

$$R_{c,t} = \gamma_0 + \gamma_1 P_c + \gamma_2 I_c + \gamma_3 F_{c,t} * P_c + \gamma_4 T_t + \sum_{i(i>4)} \gamma_i C_{i,c} + \varepsilon_{c,t} \tag{5.15}$$

$$R_{c,t} = \gamma_0 + \gamma_1 P_c + \gamma_2 I_c + \gamma_3 F_{c,t} * I_c + \gamma_4 T_t + \sum_{i(i>4)} \gamma_i C_{i,c} + \varepsilon_{c,t} \tag{5.16}$$

where P_c is the deprivation index of local community c, I_c is the Gini index of local community c, $F_{c,t}$ is the fragmentation of local information, T_t is the random-effect time variable in month t, $C_{i,c}$ is the ith control variable in community c, and ε is the error term.

Although there is no result, except one, that shows more than a 90% confidence in estimating posteriors, all the models show the same pattern: the fragmentation of local information (β_{CO}) negatively moderates the effects of socio-economic deprivation on cultural activity diversity (Table G.57, G.58, and G.59). Also, the fragmentation of local information negatively moderates the effects of socio-economic inequality in most cases (Table G.60 and G.61), particularly when Rao-Stirling index is used as a proxy for cultural activity diversity.

These patterns suggest that the fragmentation of local information might have a negative effect on the diversity of cultural activities. This implication is consistent with the findings from H3 where controlling socio-economic deprivation leads to surfacing the negative effect of the fragmentation of local information on community engagement. Because the effects of socio-economic inequality and deprivation are automatically controlled in examining the moderating effects of information fragmentation (because they are main IVs), it is possible that the pure effects of fragmentation has surfaced. Although it is not yet clear whether cultural activity diversity is beneficial or detrimental to each community, the discussion about the potentially-negative effect of information fragmentation in RQ3 still holds in the results of the H5 tests.

Overall, the results do not reject the null hypothesis of H5, but probabilistically indicate that the hypothesis might be supported. If there are more data points, this question will be able to be answered with a better statistical power.

5.5 Discussion and Limitations

5.5.1 Implications for the Fragmentation of Local Information

One of the surprising and counterintuitive findings is the effects of the fragmentation of local information in predicting community engagement. Sometimes, the fragmentation of local information is a proxy of a diversified use of information sources by residents, which is beneficial to community engagement. In other times, this construct presents a negative effect on community engagement, which follows the original meaning of it as a structural/material state that could be detrimental to people's information access.

These different behaviors indicate that the fragmentation of local information is not always the result of high inequality and deprivation in a community; rather, it is a function of the social construction of information infrastructures and the interactions between socio-economic/demographic groups. Even if an urban area is economically segregated and socially deprived, people might still enjoy sharing and using local information and participate in diverse cultural activities that are enabled and acknowledged by technological infrastructures. In such an environment, technology might have been broadly embedded in their cultural norms and practices. Although it was possible to tease out the negative effects of the fragmentation of local information in the statistical models, its effects and social meanings are still

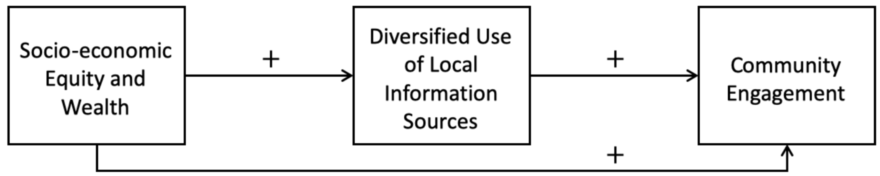

Figure 5.5: A reversed form of the theoretical model based on the empirical findings. This model is potentially possible in certain contexts.

contingent on how the socio-economic and socio-demographic groups are structured in an urban and how people enact within the socio-technical system of a community (i.e., technology-assisted local community).

When the fragmentation of local information is a proxy for the diversified use of local information sources, it is possible that the theoretical relationships between the characteristics of local communities (Figure 3.1) can be re-conceptualized. If socio-economic inequality and deprivation are reversed to socio-economic equity and wealth, the directions of the correlations are converted as well. The reversed theoretical relationships can be presented as Figure 5.5. This new model suggests that, in a particular context of urban areas, the diversified use of local information sources can positively mediate the effects of socio-economic equity and well-being on community engagement. Of course, this theoretical model is only one possible form indicated by the empirical findings. Further investigations are needed to confirm the relationships.

Despite the limitations in explicitly identifying the effect of technology penetration and the conceptual dimensions of information fragmentation, this study provides how the concept of the fragmentation of local information can be operationalized and how its measure can be used with respect to other community-level characteristics. This new construct does affect other community-level characteristics and presents some systematic patterns across different cities and times. In this sense, it is conveniently possible to say that this study provides an initial assessment of LIL theory. This also suggests that concepts about information deserts and their measures are worth further developed in the field of information science, and in a broader academic community that studies local communities and social inequality.

To further improve our understanding of the fragmentation of local information, three methodological strategies can be designed and adopted. If individual-level or high-resolution data that provide a proxy for technology penetration and adoption is combined in the analytical models, this study can be improved by characterizing each urban area based on the characteristics of the residents. This is why traditional studies on civic engagement and community characteristics have evolved in a way that combines individual-level data [Abascal and Baldassarri, 2015a, Sturgis et al., 2014, Stolle and Harell, 2013].

Another way to cope with the methodological challenges and inconsistent findings is to maximize the random sampling for target cities and information sources. If sampling bias can be reduced, it might be possible to focus solely on the materiality of local information in understanding the community-level dynamics regarding information deserts. Practically, randomized trials on these kinds of data are challenging due to a limited number of universal data sources and limitations in data collection methods. Even though, through partnerships with data providers and theoretical development of community-level processes, it would be possible to keep developing this stream of research.

Finally, simulations and predictions can be used to generate unknown variables. This study already makes use of many predicted variables and models. It is still possible to model how the data from different platforms is generated and used through training the data-generation patterns in each urban area. Although this might involve sophisticated modelings and observations, this is a practically useful way to make it possible to randomly sample cities and information sources.

5.5.2 Implications for Socio-economic Inequality and Deprivation

Similar to the concepts of the fragmentation of local information, socio-economic inequality and deprivation also connote two contradictory meanings depending on how urban areas are sampled and how people use technology in accessing local information. According to previous studies, high rates of socio-economic inequality and deprivation that are confounded with ethnoracial heterogeneity are negatively related to social trust and cohesion [Sturgis et al., 2014]. In the context of local events, however, these variables sometimes have different effects on community performance indicators such as increasing cultural activity diversity or decreasing the fragmentation of local information. Particularly in urban areas with high tech-startup culture, socio-economic inequality and deprivation play a role as a source of diversified use of technology and, thus, a source of diverse activities.

On the one hand, this might be due to the study context where community engagement, the fragmentation of local information, and cultural activity diversity are measured based on online footprints, which could bias the demographics of the subjects. On the other hand, socio-economic inequality might have to be understood as one of the factors that give rise to the diversity of socio-cultural dimensions of a community. If it is understood in combination with poverty and deprivation, it would conceptualize a dark side of economic systems that drives segregations between socio-economic classes. However, in cases where a baseline is above the poverty threshold, some variations of socio-economic status could be stimuli that could bring in more diverse perspectives and values to society.

These contradictory interpretations suggest that socio-economic features have a duality between diversity and inequality. The empirical findings imply that this duality might be shaped by people's use of technology as information sources and its embeddedness in their daily lives and information practices.

5.5.3 Technology Penetration as an Intermediary Process

Together, there are dualities or tensions between diversity and inequality (from a social-construction perspective) or diversified affordances and fragmentation (from a material perspective) existing on socio-economic features and the fragmentation of local information. It is implied by the empirical findings that these dualities are partially shaped by technology penetration or infusion in an urban area.[12] Therefore, technology penetration is an intermediary process that needs to be taken into account in examining the effects and meanings of socio-economic inequality/deprivation and the fragmentation of local information [Zmud and Ap-

[12]In the literature, the term "IT penetration" is often used in the context of hardware- and infrastructure-level implementation in a community, while "adoption" or "infusion" is used in the contexts of individual- or organization-level use of IT products. In this discussion, these terms are not distinguished.

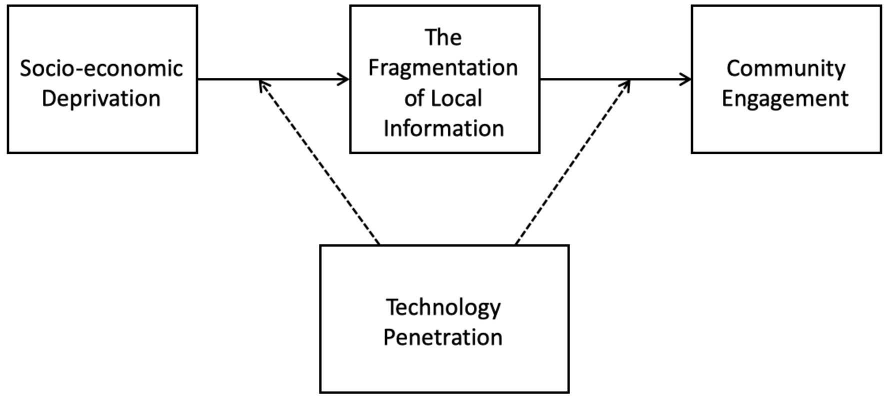

Figure 5.6: A theoretical model that can be potentially examined in future work, informed by this study. Technology penetration might moderate the effects of (1) socio-economic inequality/deprivation and (2) the fragmentation of local information.

ple, 1992, Kraemer et al., 2005]. While the two different samplings of urban areas in this study are based on tech-startup activities and population, the study does not explicitly conceptualize and operationalize technology penetration, a potential moderator in the relationship between the fragmentation of local information and community engagement, in the analytical models. As such, a theoretical model that could be potentially constructed in future studies, which is informed by the finding of this study, is depicted in Figure 5.6.

5.5.4 Implications for the Theories of Community Engagement

Theories about civic and community engagement can benefit from this study. Traditionally, studies about community heterogeneity and civic engagement in sociology, political science, and economics have focused on identifying confounded factors and intermediary processes that are related to ethnoracial compositions and people's participation in civic issues [Sturgis et al., 2011, 2014, Stolle and Harell, 2013, Stolle et al., 2008, Abascal and Baldassarri, 2015b]. While not directly related to the theoretical frameworks of these studies, recent studies have found that information that is disseminated through online platforms intervene people's civic engagement patterns such as voting [Bessi and Ferrara, 2016, Gorodnichenko et al., 2018].

In a similar vein, this study suggests that studies that focus on community engagement need to take the factors that are related to local information into account. Community engagement is affected not only by community heterogeneity and socioeconomic features, but also by the fragmentation of local information, a core part of increasing people's awareness of local information. In some sense, monitoring and managing the structure and features of local information is more critical than bot activities on Twitter because, unlike political participation, this type of community engagement shapes people's daily lives and the health of local communities, and ultimately, makes a community resilient.

In this regard, this study extends the theories of community engagement by including information-driven factors in the framework. Because the fragmentation of local information is only a part of information deserts, this is a call for researchers to extend the theoretical frameworks by studying multiple dimensions of information deserts.

5.5.5 Methodological Implications

The computational modelings of key variables are based on (1) methods developed in different fields and (2) supervised/unsupervised learning methods. The selection and development of the measures and parameters are partially tested through conceptual reasonings and benchmarks. In other words, internal validity testings and conceptual examinations are reasonably conducted in this study. Although the external validity of the computational approach to generating variables and developing measures is also partially tested through the empirical study and qualitative examinations of the output, there are many opportunities where these approaches can be improved.

The measure for the fragmentation of local information can be improved by the efforts to provide fine-grained control over the parameters for weights between balance and overlap. As briefly mentioned in Chapter 4, the current measure relies on one of the beta diversity measures developed in the field of ecology. Based on the extensive development of beta diversity measures in the field, it was possible to test the sensitivity between the measures based on the disambiguated event datasets. However, it is still desirable to develop a generalizable measure for better sensitivity analysis. Particularly in the context of local event information, the variation of "overlap" is very small while that of "balance" is very big. Because it is not clear whether this pattern would be consistent in other types of local information, it is not only conceptually challenging to define what "fragmentation" means, but also methodologically intractable to develop a universal measure. Developing a measure that allows to systematically control parameters for overlap and balance would solve this methodological issue.

The topic modeling process to measure cultural activity diversity can be improved as well. The processes to decide the number of topics and to qualitatively examine topic modeling outputs provide a baseline framework to explore the quantification of cultural activity diversity. Even if internal validity and external validity are tested through benchmarking algorithms and qualitative examinations, it is still not clear whether the topic modeling results would capture important topics well when the study scales up. For example, if gender or age balance is skewed in the training dataset, the classifications of cultural topics that are predicted from this train set could be less meaningful in characterizing the cultural activity diversity of other urban areas. Due to this reason, there need to be further explorations on how to classify topics in a way that systematically caters to the study context.

Finally, the performance of data disambiguation algorithm can be improved. Using machine learning techniques to disambiguate event datasets is not a new approach in the field of information science [Han et al., 2004]. Because of the complexity and inaccurate information in the local event datasets, however, data cleaning/curation methods and benchmarks for testing data disambiguation performance need to be enhanced. For example, Amazon M-turk was tested as an alternative to ensure the data disambiguation quality, but it turned out from a performance test that machine learning models that were trained based on manual codings outperformed Turkers' work. Also, some geo-coordinates data was found to be poorly curated in many events' location fields. These data quality and performance issues affect the precision of the measures; moreover, it will be more so if the study is designed based on a finer resolution of geographical region.

While the data disambiguation performance is reasonable for this study, it is highly possible that scaling this study up to many cities would increase the error rates as well, especially if more information sources are added. Because of this potential issue, error rate needs to be monitored using proper benchmarks in each step of data processing.

5.5.6 Implications for Neighborhood-level Studies

This study uses U.S. Census' urban area as the geographical unit of analysis, which is determined based on the survey report of average commute distance (i.e., 24 km (15 miles) to 80 km (50 miles)). This unit is smaller than the Census Metro/Micropolitan statistical area and bigger than the zip-code or city-level coverage. Although this unit is reasonably selected for meeting the purpose of this study, future studies on information deserts might have to target neighborhood as the unit of analysis. Because many types of local events are created and used at the level of neighborhood (e.g., neighborhood gatherings), there are many opportunities in studying such a study in the future.

However, studying the information dynamics of local events at the neighborhood level poses several challenges. First of all, the types of information sources that are used in a neighborhood are way more diverse than those of an urban area. Local information in a neighborhood is often disseminated through its own blogs, newspapers, mailing lists, and word-of-mouth. These kinds of information sources are highly contextualized in each neighborhood and often do not exist across different neighborhoods. As a result, these inconsistent systems through which information is communicated make it difficult to compare the local information landscapes between two or more different neighborhoods in a quantitative way.

Additionally, municipal boundaries or census-driven boundaries usually do not capture people's behavioral patterns and geographically-defined communities [Cranshaw et al., 2012]. Especially from an information landscape perspective, people might be interested in activities that occur in particular streets or regions depending on one's perceived neighborhood boundaries. This suggests that using municipal boundaries could mislead the analysis of local information landscapes. One way to deal with this challenge is to computationally identify neighborhood boundaries using people's digital footprints. For example, Cranshaw et al. [2012] used a clustering method based on the social structures of urban places (e.g., people's place check-in patterns). Also, McKenzie et al. [2018] harvested geo-located neighborhood names from CraigsList's rental postings using machine learning and identified fuzzy neighborhood boundaries in people's perceptions. Although using only a part of the online data could present demographical bias, this approach would be better in capturing the dynamically-changing neighborhood boundaries than municipal ones.

In these regards, new computational techniques, sampling methods, data collection methods need to be developed in addition to those developed in this study for the study of neighborhood-level information landscapes.

Chapter 6: Discussion and Conclusion

This dissertation has developed a theory of local information landscapes (LIL theory) and presented an empirical study to provide an initial assessment of LIL theory. The theorization process and empirical study provide implications and opportunities that help better understand the dynamics of local information landscapes.

6.1 Theoretical Contributions

Separating an information layer from the physical and technological systems have been explored to some degree among information and social science scholars [Weng et al., 2013]. LIL theory not only separates information layers from the local community, but also characterizes how information is embedded in different entities and how each component limits or extends the capability of another depending on their scales and complexity. Through this theorizing process, it was possible to examine the opportunities that were suggested at the beginning of the dissertation.

1. Local information can be provided to physical spaces, technical infrastructures, and social systems.

2. By constructing the LIL model that describes the material structure and dis-

tribution of local information at the community level, it becomes possible to understand peoples information access issues as a result of a community-level complexity in the local information landscape.

3. LIL theory can complement existing theories about information behavior, access, and contexts such as the information worlds theory and information grounds by distinguishing information provision, the structure of provided information, and information behavior/access.

4. Physical spaces, technical infrastructures, and social systems can be in different forms with varying scales and complexities that together shape the characteristics of local information.

5. Diverse manifestations of information inequality and information poverty have necessary conditions, i.e., the material states of local information landscapes. These structural forms can be conceptualized as information deserts.

Based on these opportunities and takeaways, the empirical study provides further implications for LIL theory.

6.1.1 Interplay between Technology and People

Although LIL theory provides an ontological model of how local information exists in a community using the LIL model, the conceptualization of the interplay between LIL components is still very abstract in understanding the manifestation of them. The empirical study that examines the effects of the fragmentation of local information implies how the interplay between technological infrastructure and people look. This study focused on the data collected only from three technological platforms, so did not have a chance to examine the interplay with other components of the LIL model. However, the study suggested that technology penetration, that is, people's awareness and use of information platforms, potentially gives rise to their information access, which connects to community engagement.

If viewed from the theories of community engagement, as discussed in Chapter 5, this finding can be described as the theoretical model depicted in Figure 5.6. From a LIL theory perspective, the role of technology penetration can be re-conceptualized. While technology penetration, the extent to which people use and know about technological platforms, is a community characteristic, people's awareness of local information can be understood in a material way. When people are aware of local information that is available from technological platforms, they, as a material entity, embed local information, if viewed from a LIL lens.[1] If so, "technological infrastructures" as a component of the LIL model extended the capability of "people," which is an information container, with a help of technology penetration.

An opposite scenario is also possible. If people were not aware of local events because they did not use any of the technological platforms, "technological infrastructures" did not extend the capability of people enough due to a limited technology penetration rate. This new way of viewing the interplay between technology and people provides an instance of how a LIL component extends or limits the capability

[1]People's awareness of local information can be separated from their use of them, that is, participation in local events, if viewed from a LIL lens.

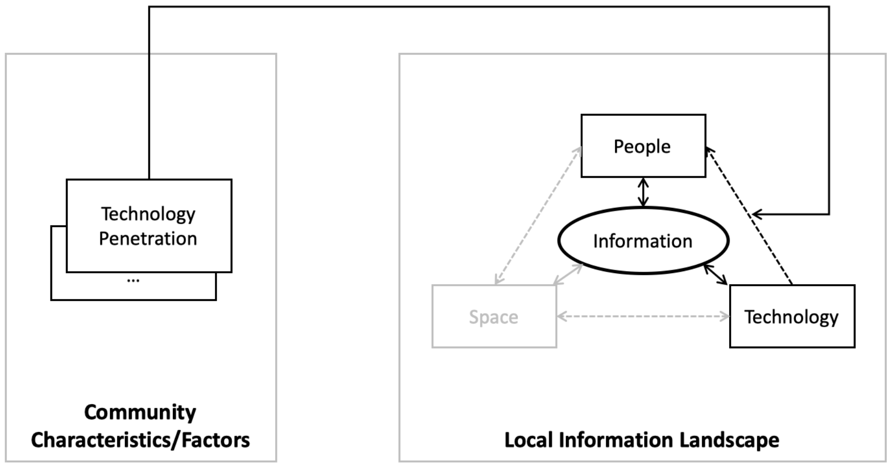

Figure 6.1: An interplay between people and technological infrastructures from a LIL perspective.

of another. Figure 6.1 shows an example of how the interplay between people and technological infrastructures look from a LIL theory perspective.

While similar to Figure 5.6, this new diagram based on LIL theory provides a new perspective on how we can understand local information landscapes. From a LIL perspective, technology penetration is not just a factor that moderates the effects of the fragmentation of local information; rather, it is an indirect way to understand how technological infrastructures and people, both as material entities that embed information, interplay with each other. Because it is practically difficult to identify the extent to which "people" contain local information, the LIL lens makes it possible to understand and operationalize the fragmentation of local information (or other dimensions of information deserts) in a holistic way. Generally speaking, information deserts can be operationalized using existing data about information sources, but can be captured and understood indirectly through identifying moderating factors that are associated with people's awareness of information.

6.1.2 Implications for Information Access and Accessibility

Because the concept of information access is characterized by the interaction between information users and the forms of how information is made available, user characteristics and environmental factors are keys to understand it [Jaeger and Burnett, 2010]. LIL theory complements the concepts and theories of information access by providing new theoretical constructs. For example, social factors that give rise to an individual's information access include institutional, organizational, and cultural forces. LIL theory suggests that the structure and feature of local information at the community level also shapes people's social aspect of information access by limiting or extending their capability to acquire local information. This material construct can shed light on the discussions about how information access is influenced not only by socially-constructed factors, but also by material forces even before information is accessed by users.

Meanwhile, LIL theory extends the concepts of information accessibility. Information accessibility has been conceptualized by mainly focusing on data format/quality, information organization, and support services [Hill, 2013]. While not explicit, these aspects of information accessibility are close to the materiality of information. By conceptualizing the community-level structure and features of local information, LIL theory indicates that information accessibility needs to include the concept of local information landscapes and information deserts, because these community-level features also characterize whether information is accessible or not. Especially when it comes to local information that people use in their daily lives (as opposed to information about books or collections), the structural and material features become critical in shaping their information practices.

Therefore, LIL theory is an initial effort to extend the discussions about both information access and accessibility, so they can be studied at the community level. Information has been often studied in particular contexts because information is understood as "collective understanding of the entity." While this conceptualization helps contextualize people's using and seeking of information, this dissertation complements the current research stream in information science from an ecological perspective by focusing on the materiality of information. Because people's information behavior and access are always influenced by ecological forces and structural features, recognizing community-level information landscapes will help understand information behaviors and practices.

6.1.3 Implications for Digital Divide

As reviewed in Chapter 1, digital divide has been categorized into first-level, second-level, and third-level digital divide [van Deursen and van Dijk, 2019, Van Deursen and Helsper, 2015]. While these characterizations provide useful ways to think about information inequality in local communities, there is a gap in understanding information inequality at the community level. From a LIL perspective, the first- and second-level divide can be understood as technology penetration and adoption, instances of community characteristics/factors in Figure 2.4, and third-level divide as community outcomes.

This indicates that the models of digital divide provide a weak understanding of information inequality itself. Even if a technology penetration rate is high in a community, everybody has adopted technological devices, and they benefit from using them, it does not mean they have access to information that is available in a community. Information infrastructures are way broader than technological infrastructures in a community, so focusing on technology and technology-driven benefits provides a partial solution and understanding of information access.

Of course, the focus of digital divide itself is on whether people have digital access to information, technology is a core part of the conceptualization. However, in order for people to have digital access to information, information needs to be provided to digital platforms and structured in advance, so users can navigate it easily. In this sense, LIL theory not only conceptualizes the gap between technology use and benefits (i.e., local information landscapes), but also theorize the information provision processes and the structure of it, which shapes people's digital access.

As such, studies about digital divide can benefit from a broader spectrum of constructs and the factors that exacerbate or are affected by the digital divide through incorporating LIL theory in the theoretical frameworks.

6.2 Practical Contributions

If the characteristics and structures of local information can be measured at the community level, there are many opportunities for practitioners to make use of the frameworks and measurements. Information deserts indicators can be monitored by policy-makers and information access professionals, as a means to manage community performance indicators such as community engagement and cultural activity diversity. In other words, once the measurements are developed in a robust manner, it becomes possible to make use of the information deserts indicators in managing local communities and in implementing information policies, which will eventually improve other community performance measures. For example, the empirical study in this dissertation shows that the fragmentation of local information can decrease people's participation in local events. By implementing policies or systems that can decrease the fragmentation indicator by 0.1%, for instance, it becomes possible to increase the local event participation rate by three people per event. From a community perspective, this change management is of a great deal in terms of increasing community members' infrastructuring process, as well as for their self-development and happiness. In this sense, this dissertation provides assessment frameworks and baseline tools for practitioners to monitor and manage local information landscapes.

6.3 Conclusion

This dissertation strives to understand information accessibility at the community level through theorization and computational methods. The goals of this dissertation, that is, (1) understanding and theorizing community-level information accessibility and (2) providing an initial assessment of LIL theory through an empirical study, are successfully achieved. Because there are still theoretical gaps and methodological challenges found in the studies, it is necessary to further develop the current theoretical frameworks and computational techniques, in order to provide practical benefits to policy-makers and practitioners. I hope this work is a starting point to think about how to connect and involve information users in understanding the ecology of information infrastructures and how to manage unevenly fluctuating pieces of information in the community where we live.

Appendix A: List of Target Cities and Numbers of Events

City	# Meetup	# Yelp	#Eventful	Data Period
Austin, TX	52,726	1,733	30,424	Jan. 2017 - Aug. 2018
Baltimore, MD	29,007	833	29,866	Jan. 2017 - Aug. 2018
Boston, MA	52,768	2,661	58,687	Jan. 2017 - Aug. 2018
Charlotte, NC	3,352	128	5,603	June 2018 - Aug. 2018
Chicago, IL	7,617	484	22,569	June 2018 - Aug. 2018
Columbus, OH	1,934	129	5,548	June 2018 - Aug. 2018
Dallas, TX	7,127	122	11,747	June 2018 - Aug. 2018
Denver, CO	79,589	1,458	34,707	Jan. 2017 - Aug. 2018
Detroit, MI	2,704	119	7,652	June 2018 - Aug. 2018
Durham, NC	39,457	233	15,808	Jan. 2017 - Aug. 2018
El Paso, TX	318	27	1,277	June 2018 - Aug. 2018
Fort Worth, TX	2,279	56	4,147	June 2018 - Aug. 2018
Houston, TX	4,702	274	9,753	June 2018 - Aug. 2018
Indianapolis, IN	1,641	182	6,656	June 2018 - Aug. 2018
Jacksonville, FL	1,423	72	1,898	June 2018 - Aug. 2018

Los Angeles, CA	134,355	2,818	129,568	Jan. 2017 - Aug. 2018
Memphis TN	488	31	2,168	June 2018 - Aug. 2018
Nashville, TN	2,030	240	5,280	June 2018 - Aug. 2018
New York, NY	170,450	7,301	241,256	Jan. 2017 - Aug. 2018
Philadelphia, PA	54,501	2,052	50,960	Jan. 2017 - Aug. 2018
Phoenix, AZ	7,422	77	9,354	June 2018 - Aug. 2018
Pittsburgh, PA	13,228	1,116	14,179	Jan. 2017 - Aug. 2018
Raleigh, NC	38,410	434	13,951	Jan. 2017 - Aug. 2018
San Antonio, TX	2,425	376	3,272	June 2018 - Aug. 2018
San Diego, CA	55,307	2,317	41,822	Jan. 2017 - Aug. 2018
San Francisco, CA	93,522	5,554	81,202	Jan. 2017 - Aug. 2018
San Jose, CA	70,118	1,272	46,844	Jan. 2017 - Aug. 2018
Seattle, WA	9,069	179	19,245	June 2018 - Aug. 2018
Washington, DC	105,023	2,452	99,701	Jan. 2017 - Aug. 2018

Table A.1: The list of target cities and numbers of events per each city.

Rank	State	Number of Cities
1	California (CA)	4
2	North Carolina (NC)	2
2	Pennsylvania (PA)	2
3	Colorado (CO)	1
3	Washington D.C.	1
3	Massachusetts (MA)	1
3	Maryland (MD)	1
3	New York (NY)	1
3	Illinois (IL)	1
3	Texas (TX)	1

Table A.2: The number of cities per state in a descending order for the 14-city dataset.

Rank	State	Number of Cities
1	Texas (TX)	5
2	California (CA)	4
3	North Carolina (NC)	3
4	Pennsylvania (PA)	2
4	Tennessee (TN)	2
5	Colorado (CO)	1
5	Washington D.C.	1
5	Florida (FL)	1
5	Illinois (IL)	1
5	Massachusetts (MA)	1
5	Maryland (MD)	1
5	Michigan (MI)	1
5	New York (NY)	1
5	Ohio (OH)	1
5	Washington (WA)	1

Table A.3: The number of cities per state in a descending order for the 28-city dataset.

Appendix B: Topic Modeling Validity Tests: Perplexity Plots and LDAvis

Figure B.1, Figure B.2, and Figure B.3 show the changes of perplexity scores when α and β priors change. Because the lower perplexity score indicates a better clustering of topics, $\alpha = 0.2$ and $\beta = 0.7$ for $N_T = 48$, $\alpha = 0.2$ and $\beta = 0.7$ when $N_T = 48$ and 60, and $\alpha = 0.1$ and $\beta = 0.8$ when $N_T = 90$ present the best internal validity in the topic modeling. Through qualitative examinations using the LDAvis interface shown in Figure B.4, these candidates are examined qualitatively. Because it is ideal if the distribution of potential participants across the events with the generated topics can vary depending on demographic and socio-economic groups, as reviewed in Section 5.1.2, variations of topics within a high-level topic are of interests in the qualitative examination process.

For example, the "music" category, a high-level topic that can be found in the information sources' own categories, does not give rise to the variations between participants from different socio-demographic or socio-economic groups. However, if the "music" category is divided into multiple categories such as "classical music" and "live guitars in the bar," participants in these two different categories of events might vary based on their socio-demographic and socio-economic groups.

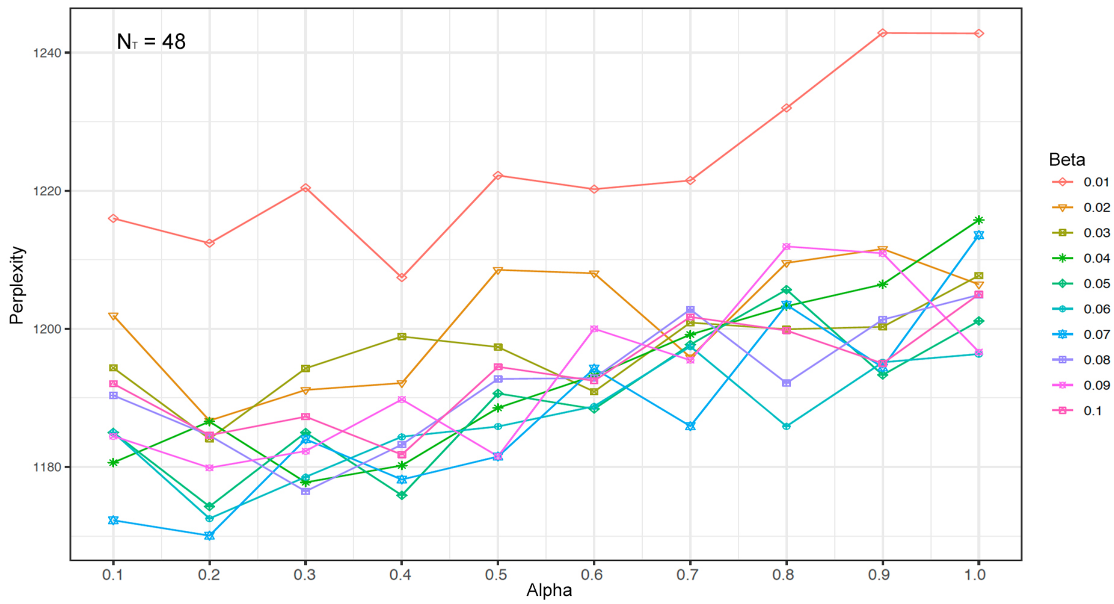

Figure B.1: Perplexity changes when α and β change (N_T=48).

In the topic modeling results of the 48-topic solution ($\alpha = 0.2$ and $\beta = 0.7$), there are about five music-related topics when considering only major words. These five topics are characterized as (1) Jazz, Bruce, Rock, Band, Punk, etc., (2) concerts, performance, opera, symphony, piano, choir, violin, etc., (3) party, night-club, disco, hip, lounge, 80s/90s, night-life, DJ, rooftop, (4) festival, summer, annual, pop, celebration, Christmas, food, seasonal, etc., and (5) trivia, pub, karaoke, bar, happy-hour, 20s, singles, etc. These categorizations would reasonably vary participants from different socio-economic/demographic groups. Because the 48-topic solution presents the most reasonable classifications of event topics (i.e., topics not overlapping too much with each other while each topic is not too general), N_T is determined as 48.

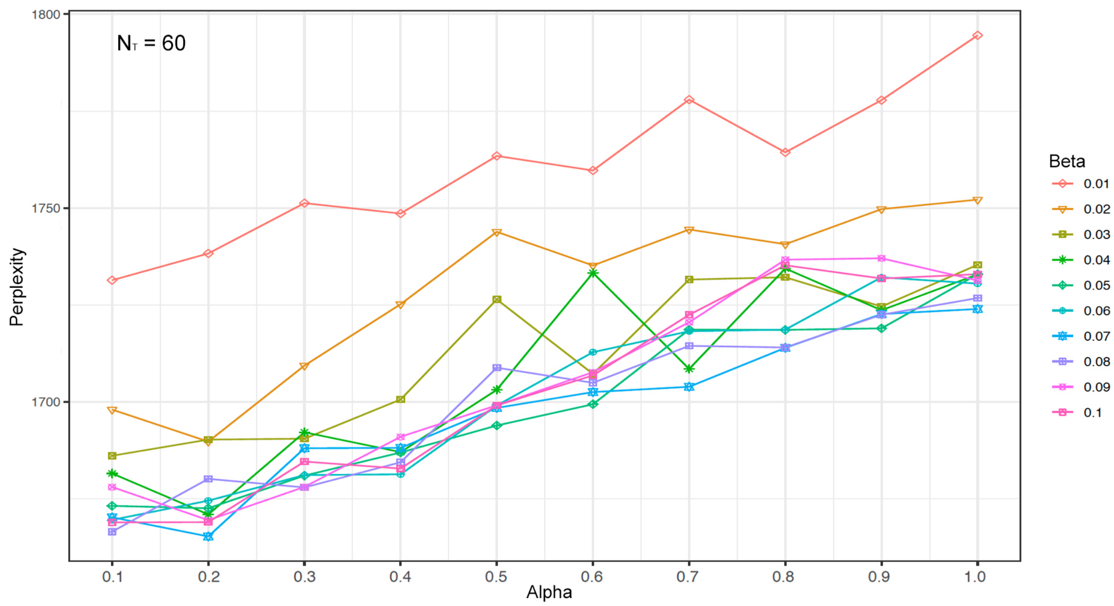

Figure B.2: Perplexity changes when α and β change (N_T=60).

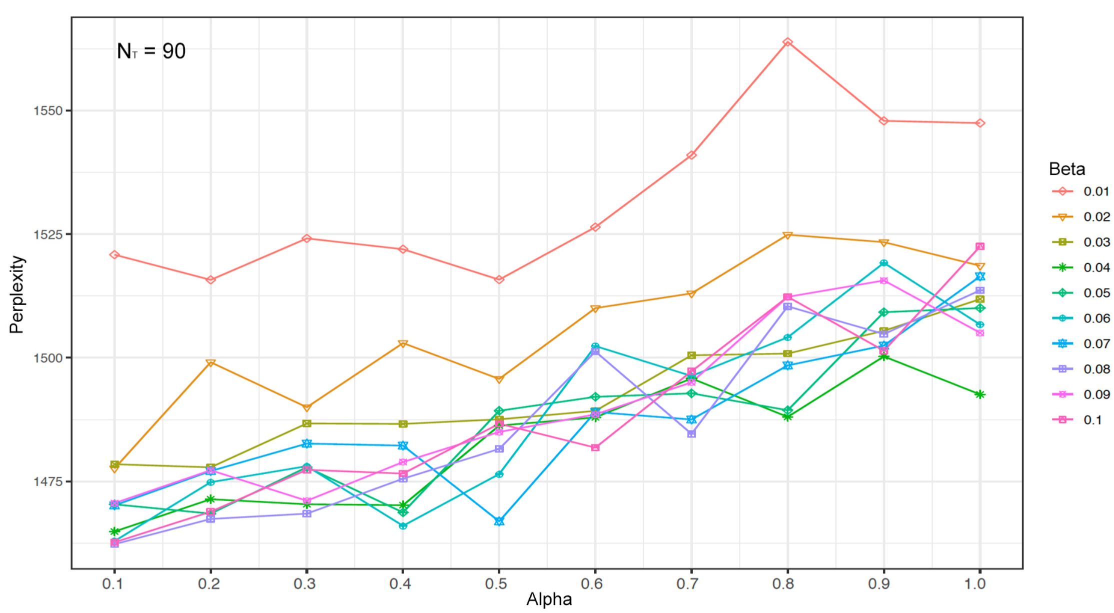

Figure B.3: Perplexity changes when α and β change (N_T=90).

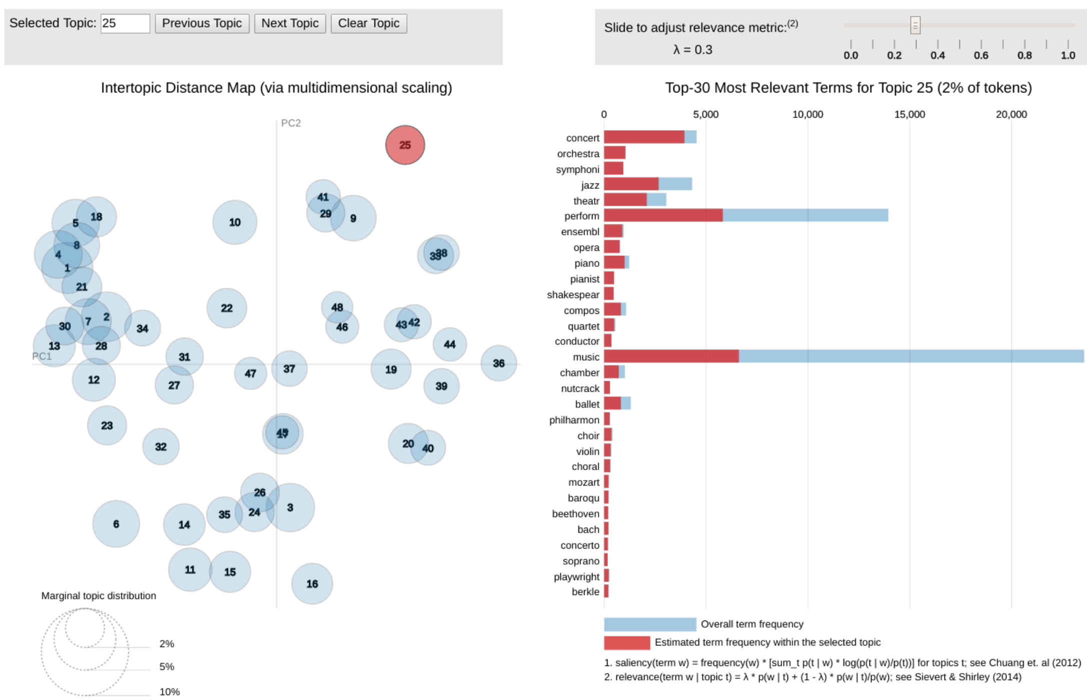

Figure B.4: Screenshot of the qualitative examination tool for LDA results (LDAvis) when N_T=48. Topic 25 shows a music-related activity category that is mostly classical.

Appendix C: Topics and Their Word Distributions Generated from LDA

Topic	Word Distribution
1	[λ = 1]: hour (0.053), night (0.051), happi (0.041), drink (0.031), tuesday (0.026), even (0.022), bar (0.021), at (0.021), week (0.019), wednesday (0.018), friday (0.018), thursday (0.015), monday (0.015), food (0.012), special (0.011), super (0.011), fun (0.011), friend (0.011), class (0.01), bowl (0.01) [λ = 0.3]: happi (0.041), clay (0.008), potteri (0.007), mic (0.008), flex (0.005), ceram (0.006), bowl (0.01), wheel (0.007), hour (0.053), climb (0.01), tuesday (0.026), throw (0.008), belay (0.003), tavern (0.005), night (0.051), taco (0.006), drink (0.031), pizza (0.009), pub (0.007), super (0.011)
2	[λ = 1]: race (0.039), run (0.038), annual (0.024), walk (0.024), celebr (0.023), 5k (0.02), donat (0.015), rais (0.013), benefit (0.012), nation (0.011), support (0.011), chariti (0.011), complet (0.011), choos (0.011), virtual (0.011), 10k (0.01), fundrais (0.01), veteran (0.01), for (0.01), ship (0.01) [λ = 0.3]: race (0.039), 5k (0.02), chariti (0.011), medal (0.01), 10k (0.01), ship (0.01), fundrais (0.01), veteran (0.01), cancer (0.007), bib (0.003), marathon (0.005), auction (0.005), run (0.038), sweater (0.003), rais (0.013), ugli (0.003), annual (0.024), proceed (0.007), jog (0.002), dash (0.002)
3	[λ = 1]: data (0.025), blockchain (0.022), develop (0.018), technolog (0.016), train (0.012), grant (0.012), applic (0.01), cloud (0.008), bitcoin (0.008), build (0.007), code (0.007), comput (0.007), softwar (0.006), secur (0.006), write (0.006), web (0.006), career (0.006), provid (0.006), talk (0.005), machin (0.005)

	[λ = 0.3]: blockchain (0.022), data (0.025), bitcoin (0.008), grant (0.012), technolog (0.016), aw (0.004), ethereum (0.003), cloud (0.008), javascript (0.003), omni212 (0.003), python (0.005), cryptocurr (0.004), hyperledg (0.002), machin (0.005), applic (0.01), sql (0.002), io (0.002), azur (0.002), android (0.002), tableau (0.002)
4	[λ = 1]: saturday (0.053), pm (0.051), sunday (0.04), friday (0.027), juli (0.023), june (0.022), 00pm (0.021), august (0.02), april (0.018), 30pm (0.018), march (0.017), thursday (0.016), januari (0.016), octob (0.013), septemb (0.013), novemb (0.013), februari (0.011), annual (0.01), decemb (0.01), weekend (0.009) [λ = 0.3]: saturday (0.053), sunday (0.04), june (0.022), august (0.02), juli (0.023), april (0.018), januari (0.016), pm (0.051), septemb (0.013), march (0.017), 00pm (0.021), octob (0.013), novemb (0.013), sat (0.006), 28th (0.004), 11am (0.005), mon (0.004), februari (0.011), feb (0.005), nov (0.005)
5	[λ = 1]: ticket (0.06), door (0.021), bar (0.021), parti (0.019), guest (0.016), drink (0.015), special (0.014), vip (0.013), admiss (0.013), seat (0.012), purchas (0.012), night (0.012), reserv (0.011), tabl (0.011), includ (0.01), loung (0.01), dress (0.01), dj (0.01), saturday (0.009), bottl (0.008) [λ = 0.3]: ticket (0.06), vip (0.013), loung (0.01), rooftop (0.007), crawl (0.008), hookah (0.003), door (0.021), admiss (0.013), bar (0.021), dress (0.01), guestlist (0.001), champagn (0.004), 11pm (0.003), bottl (0.008), upscal (0.002), entri (0.008), attir (0.003), wristband (0.001), 10pm (0.005), dj (0.01)
6	[λ = 1]: ride (0.072), run (0.049), mile (0.024), bike (0.022), rout (0.016), pace (0.016), road (0.014), rider (0.01), park (0.008), trail (0.008), stop (0.007), water (0.007), morn (0.007), skate (0.007), runner (0.006), distanc (0.006), hill (0.006), minut (0.006), loop (0.006), light (0.006) [λ = 0.3]: ride (0.072), bike (0.022), run (0.049), rout (0.016), rider (0.01), skate (0.007), helmet (0.006), pace (0.016), mile (0.024), bicycl (0.004), runner (0.006), tube (0.003), mph (0.004), road (0.014), ski (0.002), rink (0.001), loop (0.006), distanc (0.006), cycl (0.005), slower (0.001)

7	[$\lambda = 1$]: age (0.046), famili (0.046), children (0.036), kid (0.034), adult (0.029), school (0.024), librari (0.021), fun (0.021), activ (0.018), child (0.018), parent (0.017), program (0.016), friend (0.014), team (0.012), stori (0.011), enjoy (0.011), reserv (0.011), boston (0.009), teen (0.009), babi (0.008) [$\lambda = 0.3$]: kid (0.034), children (0.036), child (0.018), adult (0.029), parent (0.017), famili (0.046), age (0.046), librari (0.021), teen (0.009), babi (0.008), branch (0.005), preschool (0.004), toddler (0.003), mom (0.005), school (0.024), lego (0.002), cerritotrivia (0.002), storytim (0.005), zoo (0.004), kindergarten (0.001)
8	[$\lambda = 1$]: medit (0.057), yoga (0.051), practic (0.031), class (0.029), mind (0.027), bodi (0.013), experi (0.012), breath (0.011), mat (0.01), session (0.01), guid (0.009), relax (0.009), stress (0.008), donat (0.008), level (0.008), peac (0.007), life (0.007), flow (0.007), minut (0.006), teach (0.006) [$\lambda = 0.3$]: yoga (0.051), medit (0.057), breath (0.011), mat (0.01), mind (0.027), buddhist (0.004), chi (0.004), practic (0.031), tai (0.004), buddhism (0.003), stress (0.008), dharma (0.002), vinyasa (0.002), calm (0.003), zen (0.002), retreat (0.003), postur (0.002), buddha (0.002), cushion (0.002), peac (0.007)
9	[$\lambda = 1$]: parti (0.07), night (0.043), dj (0.02), music (0.02), danc (0.017), friday (0.015), date (0.013), hop (0.012), hip (0.01), love (0.01), live (0.01), host (0.01), club (0.01), singl (0.009), at (0.009), enjoy (0.009), saturday (0.009), speed (0.009), celebr (0.008), male (0.008) [$\lambda = 0.3$]: parti (0.07), dj (0.02), hop (0.012), male (0.008), hip (0.01), disco (0.003), 90s (0.003), hottest (0.006), 80s (0.003), bash (0.004), speed (0.009), birthday (0.007), night (0.043), glow (0.003), booti (0.002), djs (0.004), cow (0.002), whisler (0.001), scorecard (0.001), claddagh (0.001)
10	[$\lambda = 1$]: scienc (0.022), health (0.02), medic (0.011), research (0.01), dr (0.01), educ (0.009), care (0.009), system (0.009), clinic (0.009), univers (0.008), program (0.007), includ (0.006), seminar (0.006), confer (0.006), technolog (0.006), center (0.006), inform (0.005), mental (0.005), provid (0.005), medicin (0.005)

	[$\lambda = 0.3$]: scienc (0.022), clinic (0.009), medic (0.011), health (0.02), nurs (0.004), treatment (0.004), patient (0.004), medicin (0.005), scientif (0.002), research (0.01), healthcar (0.004), diseas (0.003), disord (0.002), surgeri (0.002), dr (0.01), stem (0.002), therapi (0.003), phd (0.002), seminar (0.006), regulatori (0.002)
11	[$\lambda = 1$]: book (0.074), read (0.026), author (0.017), stori (0.014), languag (0.013), discuss (0.013), french (0.012), english (0.012), convers (0.009), japanes (0.008), club (0.008), mother (0.007), love (0.007), and (0.007), cultur (0.006), studi (0.006), by (0.005), with (0.005), american (0.005), sign (0.005) [$\lambda = 0.3$]: book (0.074), french (0.012), japanes (0.008), author (0.017), english (0.012), read (0.026), languag (0.013), korean (0.003), du (0.003), mother (0.007), literari (0.003), fran (0.002), reader (0.005), german (0.004), bestsel (0.002), fiction (0.004), des (0.001), pari (0.002), mandarin (0.002), ash (0.002)
12	[$\lambda = 1$]: dinner (0.027), food (0.019), restaur (0.015), chef (0.014), cook (0.013), menu (0.012), brunch (0.011), tea (0.008), eat (0.008), dish (0.008), vegan (0.008), meal (0.008), kitchen (0.007), enjoy (0.007), chees (0.007), includ (0.007), delici (0.007), serv (0.007), fresh (0.007), ice (0.006) [$\lambda = 0.3$]: chef (0.014), cook (0.013), dinner (0.027), dish (0.008), vegan (0.008), chicken (0.006), menu (0.012), salad (0.005), recip (0.005), dessert (0.006), bake (0.004), cream (0.005), sandwich (0.004), cooki (0.004), chees (0.007), kitchen (0.007), brunch (0.011), meat (0.003), chocol (0.006), cuisin (0.004)
13	[$\lambda = 1$]: fit (0.035), bodi (0.019), health (0.019), workout (0.015), weight (0.014), feel (0.014), goal (0.011), well (0.011), exercis (0.009), challeng (0.009), life (0.008), strength (0.007), person (0.007), peopl (0.007), healthi (0.007), loss (0.007), success (0.007), support (0.006), result (0.006), energi (0.006) [$\lambda = 0.3$]: fit (0.035), weight (0.014), workout (0.015), well (0.011), nutrit (0.006), strength (0.007), sheryl (0.003), health (0.019), bodi (0.019), loss (0.007), metabol (0.002), habit (0.003), healthi (0.007), muscl (0.004), fat (0.003), lose (0.004), pain (0.005), diabet (0.002), cardio (0.003), keto (0.001)

14	[$\lambda = 1$]: danc (0.113), class (0.062), lesson (0.024), salsa (0.024), beginn (0.017), pm (0.015), partner (0.013), swing (0.013), fun (0.012), social (0.011), week (0.011), dancer (0.011), level (0.01), bachata (0.009), step (0.009), music (0.009), intermedi (0.008), month (0.007), tango (0.007), floor (0.007) [$\lambda = 0.3$]: danc (0.113), salsa (0.024), lesson (0.024), swing (0.013), bachata (0.009), tango (0.007), dancer (0.011), cha (0.005), kizomba (0.005), ballroom (0.005), zumba (0.004), argentin (0.003), class (0.062), merengu (0.002), waltz (0.002), beginn (0.017), beg (0.002), choreographi (0.002), u2 (0.002), latin (0.005)
15	[$\lambda = 1$]: de (0.064), la (0.056), el (0.03), los (0.029), angel (0.02), en (0.018), spanish (0.015), paso (0.011), con (0.011), del (0.01), para (0.009), las (0.007), es (0.007), est (0.006), se (0.006), mexican (0.005), mexico (0.005), por (0.005), al (0.005), ol (0.005) [$\lambda = 0.3$]: de (0.064), la (0.056), el (0.03), en (0.018), los (0.029), spanish (0.015), del (0.01), para (0.009), con (0.011), paso (0.011), es (0.007), se (0.006), por (0.005), una (0.005), las (0.007), ol (0.005), cinco (0.004), su (0.004), espa (0.004), mayo (0.004)
16	[$\lambda = 1$]: busi (0.102), network (0.055), market (0.024), profession (0.02), onlin (0.013), grow (0.013), referr (0.012), connect (0.012), entrepreneur (0.012), build (0.011), sale (0.01), success (0.008), compani (0.008), owner (0.008), industri (0.008), card (0.008), opportun (0.008), brand (0.007), lead (0.007), relationship (0.006) [$\lambda = 0.3$]: busi (0.102), network (0.055), referr (0.012), entrepreneur (0.012), market (0.024), pitch (0.005), grow (0.013), bni (0.003), linkedin (0.003), profession (0.02), brand (0.007), owner (0.008), mastermind (0.002), advertis (0.003), silicon (0.003), meetaway (0.001), entrepreneuri (0.001), zarlun (0.001), founder (0.006), profess (0.002)
17	[$\lambda = 1$]: manag (0.027), project (0.018), design (0.016), product (0.016), develop (0.014), team (0.013), understand (0.011), process (0.009), plan (0.009), scrum (0.008), improv (0.008), custom (0.008), experi (0.007), agil (0.007), effect (0.007), organ (0.007), particip (0.007), practic (0.007), train (0.007), program (0.006)

	[λ = 0.3]: salesforc (0.003), manag (0.027), scrum (0.008), trainerkart (0.002), ux (0.002), product (0.016), cyber (0.002), design (0.016), deliveri (0.003), organiz (0.001), role (0.005), workplac (0.002), agil (0.007), effect (0.007), project (0.018), mangat (0.001), assess (0.002), kanban (0.001), smc (0.001), understand (0.011)
18	[λ = 1]: write (0.031), share (0.013), minut (0.013), peopl (0.012), check (0.011), writer (0.01), read (0.01), meetup (0.009), week (0.009), lead (0.009), session (0.009), break (0.008), talk (0.008), space (0.007), stori (0.007), idea (0.007), question (0.007), page (0.007), word (0.007), critiqu (0.006) [λ = 0.3]: write (0.031), critiqu (0.006), ittl (0.003), poetri (0.005), blog (0.004), writer (0.01), atheist (0.002), itakethelead (0.002), shut (0.002), poem (0.002), skeptic (0.001), nlp (0.002), submiss (0.002), cowork (0.002), tab (0.003), regiment (0.001), divorc (0.002), decker (0.001), stuffi (0.001), pen (0.003)
19	[λ = 1]: hike (0.033), park (0.03), trail (0.022), water (0.021), walk (0.018), mile (0.014), lake (0.012), weather (0.009), boat (0.008), river (0.008), paddl (0.008), lot (0.008), road (0.007), creek (0.007), kayak (0.006), trip (0.006), wear (0.006), mask (0.006), rain (0.005), drive (0.005) [λ = 0.3]: hike (0.033), trail (0.022), paddl (0.008), kayak (0.006), boat (0.008), lake (0.012), water (0.021), creek (0.007), river (0.008), carpool (0.003), hiker (0.002), trailhead (0.003), ridg (0.003), weather (0.009), mile (0.014), backpack (0.002), canyon (0.002), sunscreen (0.002), leash (0.002), cano (0.001)
20	[λ = 1]: love (0.031), life (0.024), attract (0.017), date (0.017), qualiti (0.016), women (0.015), live (0.015), relationship (0.014), readi (0.011), woman (0.011), step (0.01), irresist (0.009), amaz (0.008), power (0.008), heart (0.008), past (0.008), feel (0.008), coach (0.007), week (0.007), simpl (0.007) [λ = 0.3]: attract (0.017), irresist (0.009), woman (0.011), emyrald (0.004), magnet (0.004), manifest (0.005), soulmat (0.003), love (0.031), relationship (0.014), qualiti (0.016), masterclass (0.005), happili (0.002), sex (0.004), fear (0.005), wow (0.002), stuck (0.003), life (0.024), date (0.017), unstuck (0.001), guy (0.005)

21	[$\lambda = 1$]: univers (0.013), american (0.01), nation (0.01), world (0.009), public (0.007), school (0.007), cultur (0.007), director (0.007), discuss (0.007), law (0.007), women (0.006), intern (0.006), histori (0.006), presid (0.006), america (0.006), polit (0.005), global (0.005), black (0.005), program (0.005), societi (0.005) [$\lambda = 0.3$]: polit (0.005), presid (0.006), univers (0.013), immigr (0.002), law (0.007), justic (0.003), african (0.004), econom (0.003), elect (0.002), professor (0.003), violenc (0.002), civil (0.002), harvard (0.002), vice (0.002), govern (0.004), panel (0.004), trump (0.001), right (0.003), american (0.01), activist (0.001)
22	[$\lambda = 1$]: train (0.064), certif (0.038), exam (0.03), pmp (0.028), manag (0.028), project (0.021), classroom (0.014), cours (0.014), certifi (0.013), contact (0.012), pmi (0.012), profession (0.011), agil (0.009), itil (0.008), support (0.008), pass (0.008), lean (0.008), sigma (0.008), belt (0.008), acp (0.007) [$\lambda = 0.3$]: certif (0.038), exam (0.03), pmp (0.028), train (0.064), pmi (0.012), classroom (0.014), itil (0.008), sigma (0.008), acp (0.007), cours (0.014), belt (0.008), edukla (0.006), lean (0.008), certifi (0.013), capm (0.005), csm (0.004), pmbok (0.004), simul (0.005), icertglob (0.004), icert (0.003)
23	[$\lambda = 1$]: heal (0.022), life (0.018), energi (0.015), love (0.013), experi (0.013), spiritu (0.013), share (0.009), reiki (0.009), connect (0.009), live (0.009), circl (0.008), spirit (0.008), god (0.007), journey (0.007), power (0.007), feel (0.007), person (0.006), heart (0.006), light (0.006), world (0.006) [$\lambda = 0.3$]: heal (0.022), reiki (0.009), spiritu (0.013), god (0.007), bibl (0.005), sacr (0.004), energi (0.015), divin (0.005), intuit (0.004), psychic (0.003), bless (0.004), crystal (0.005), prayer (0.003), shaman (0.003), circl (0.008), tarot (0.002), worship (0.002), chakra (0.003), healer (0.002), christ (0.002)
24	[$\lambda = 1$]: art (0.077), artist (0.032), paint (0.025), exhibit (0.022), galleri (0.017), design (0.014), museum (0.014), creat (0.014), studio (0.012), color (0.011), creativ (0.011), photograph (0.01), photo (0.01), collect (0.009), photographi (0.009), draw (0.009), light (0.008), explor (0.008), imag (0.007), fashion (0.007)

	[λ = 0.3]: art (0.077), paint (0.025), exhibit (0.022), galleri (0.017), artist (0.032), photographi (0.009), photograph (0.01), color (0.011), imag (0.007), portrait (0.004), draw (0.009), museum (0.014), canva (0.003), artwork (0.003), jewelri (0.003), studio (0.012), photo (0.01), recept (0.005), camera (0.005), fashion (0.007)
25	[λ = 1]: tournament (0.027), win (0.025), prize (0.025), bonus (0.02), poker (0.02), chip (0.018), team (0.017), card (0.017), receiv (0.016), cash (0.015), game (0.014), club (0.012), beat (0.012), bad (0.012), facebook (0.011), bar (0.011), gift (0.011), play (0.011), deal (0.01), onlin (0.009) [λ = 0.3]: tournament (0.027), poker (0.02), chip (0.018), bonus (0.02), prize (0.025), win (0.025), dealer (0.008), bingo (0.005), beat (0.012), bad (0.012), koala (0.004), token (0.004), bounti (0.004), contest (0.006), 2k (0.003), raffl (0.007), bpo (0.002), cash (0.015), bbpc (0.002), stack (0.005)
26	[λ = 1]: play (0.046), player (0.03), game (0.028), rsvp (0.02), court (0.019), level (0.018), tenni (0.016), team (0.015), meetup (0.015), ball (0.012), field (0.011), fun (0.009), skill (0.009), soccer (0.008), volleybal (0.008), cancel (0.007), organ (0.007), list (0.006), peopl (0.006), leagu (0.006) [λ = 0.3]: court (0.019), tenni (0.016), player (0.03), volleybal (0.008), ball (0.012), soccer (0.008), play (0.046), pickup (0.004), picklebal (0.002), field (0.011), cleat (0.002), turf (0.002), frisbe (0.002), sand (0.003), indoor (0.005), doubl (0.005), leagu (0.006), softbal (0.001), waitlist (0.002), foul (0.001)
27	[λ = 1]: webinar (0.021), busi (0.017), team (0.016), move (0.014), step (0.013), transform (0.011), chang (0.011), coach (0.011), employe (0.01), leadership (0.01), creat (0.01), regist (0.01), strategi (0.009), to (0.009), offer (0.008), inform (0.008), peopl (0.008), life (0.008), pay (0.008), perform (0.008) [λ = 0.3]: webinar (0.021), heather (0.006), employe (0.01), trade (0.007), revenu (0.004), williamson (0.002), tangibl (0.002), carolyn (0.002), forex (0.002), avp (0.002), transform (0.011), transformationgroupllc (0.002), leadership (0.01), owe (0.002), franci (0.002), legalshield (0.001), harv (0.001), ten (0.005), move (0.014), task (0.004)
28	[λ = 1]: game (0.126), play (0.06), night (0.032), board (0.027), food (0.019), trivia (0.017), player (0.011), servsaf (0.011), card (0.01), peopl (0.009), tabl (0.009), fun (0.009), golf (0.008), drink (0.007), feel (0.007), host (0.005), week (0.005), arriv (0.005), collect (0.005), meetup (0.005)

	[$\lambda = 0.3$]: game (0.126), trivia (0.017), board (0.027), servsaf (0.011), golf (0.008), play (0.06), gamer (0.003), chess (0.002), proctor (0.002), euchr (0.002), tabletop (0.002), mah (0.002), boardgam (0.001), catan (0.001), werewolf (0.001), foodservic (0.001), jongg (0.001), whaddaya-know (0.001), buster (0.001), euro (0.001)
29	[$\lambda = 1$]: nashvill (0.02), texa (0.018), houston (0.017), north (0.016), dalla (0.015), detroit (0.014), at (0.013), carolina (0.011), star (0.01), vs (0.01), citi (0.01), nc (0.01), tx (0.009), fan (0.008), basketbal (0.008), charlott (0.007), girl (0.007), red (0.007), raleigh (0.007), harri (0.007) [$\lambda = 0.3$]: nashvill (0.02), detroit (0.014), vs (0.01), carolina (0.011), dalla (0.015), potter (0.005), texa (0.018), harri (0.007), basketbal (0.008), houston (0.017), basebal (0.004), nc (0.01), durham (0.006), stadium (0.003), fli (0.006), tiger (0.002), derbi (0.002), battl (0.005), bee (0.003), fan (0.008)
30	[$\lambda = 1$]: communiti (0.057), volunt (0.03), support (0.021), food (0.018), women (0.016), sponsor (0.015), provid (0.015), organ (0.012), local (0.011), donat (0.011), servic (0.01), age (0.01), famili (0.01), program (0.009), live (0.009), serv (0.009), center (0.008), mission (0.008), opportun (0.008), click (0.008) [$\lambda = 0.3$]: volunt (0.03), communiti (0.057), homeless (0.004), sponsor (0.015), queer (0.002), shelter (0.004), nonprofit (0.004), mission (0.008), baltimor (0.006), lgbtq (0.002), pantri (0.001), kink (0.001), youth (0.006), clean (0.005), transgend (0.001), compost (0.001), women (0.016), outreach (0.002), civic (0.002), support (0.021)
31	[$\lambda = 1$]: particip (0.028), organ (0.015), activ (0.014), safeti (0.014), sign (0.013), respons (0.012), agre (0.012), person (0.012), risk (0.011), includ (0.01), releas (0.01), injuri (0.009), waiver (0.009), liabil (0.009), attend (0.008), water (0.006), understand (0.006), limit (0.006), legal (0.006), safe (0.006) [$\lambda = 0.3$]: agre (0.012), injuri (0.009), liabil (0.009), damag (0.005), safeti (0.014), cannabi (0.005), claim (0.006), risk (0.011), waiver (0.009), particip (0.028), agreement (0.003), aris (0.003), respons (0.012), harmless (0.002), hazard (0.002), consent (0.002), dive (0.005), acknowledg (0.003), waiv (0.002), marijuana (0.001)
32	[$\lambda = 1$]: music (0.021), perform (0.017), concert (0.015), award (0.011), john (0.01), featur (0.01), artist (0.009), king (0.007), michael (0.006), david (0.006), orchestra (0.006), includ (0.006), symphoni (0.005), win (0.005), piano (0.005), toni (0.005), compos (0.005), classic (0.004), william (0.004), hall (0.004)

	[λ = 0.3]: orchestra (0.006), symphoni (0.005), concert (0.015), piano (0.005), jr (0.004), martin (0.004), compos (0.005), peter (0.004), john (0.01), conductor (0.002), pianist (0.002), michael (0.006), chamber (0.004), opera (0.003), philharmon (0.002), quartet (0.002), luther (0.002), award (0.011), tim (0.002), choir (0.002)
33	[λ = 1]: meetup (0.053), peopl (0.032), rsvp (0.03), friend (0.028), social (0.024), fun (0.022), attend (0.02), drink (0.014), singl (0.013), facebook (0.012), month (0.011), post (0.01), feel (0.009), plan (0.008), mask (0.008), link (0.008), food (0.008), page (0.008), invit (0.008), host (0.007) [λ = 0.3]: meetup (0.053), rsvp (0.03), cuddl (0.003), social (0.024), zwift (0.002), prepay (0.002), singl (0.013), friend (0.028), mixer (0.005), mingl (0.006), peopl (0.032), 30s (0.001), facebook (0.012), breaker (0.002), 20s (0.001), uber (0.002), hangout (0.002), genentech (0.001), grs (0.001), 40s (0.001)
34	[λ = 1]: music (0.041), band (0.029), song (0.018), jazz (0.014), rock (0.014), blue (0.013), play (0.012), record (0.012), album (0.011), perform (0.011), guitar (0.01), live (0.01), sound (0.009), releas (0.009), sing (0.009), soul (0.009), jam (0.008), musician (0.007), singer (0.006), songwrit (0.006) [λ = 0.3]: band (0.029), album (0.011), song (0.018), guitar (0.01), jazz (0.014), jam (0.008), songwrit (0.006), music (0.041), singer (0.006), vocal (0.005), blue (0.013), rock (0.014), acoust (0.004), bass (0.004), musician (0.007), guitarist (0.003), sing (0.009), instrument (0.005), record (0.012), chord (0.002)
35	[λ = 1]: tour (0.065), garden (0.018), walk (0.015), histor (0.015), citi (0.013), museum (0.013), histori (0.012), santa (0.011), locat (0.011), indianapoli (0.01), plant (0.009), explor (0.009), beach (0.008), visit (0.008), west (0.008), includ (0.008), hous (0.008), fort (0.008), center (0.008), district (0.007) [λ = 0.3]: tour (0.065), garden (0.018), santa (0.011), histor (0.015), fort (0.008), indianapoli (0.01), plant (0.009), marion (0.003), ref (0.003), flatiron (0.002), ocean (0.004), landmark (0.003), monument (0.002), indiana (0.004), histori (0.012), tree (0.006), district (0.007), docent (0.002), clara (0.002), island (0.006)

36	[$\lambda = 1$]: comedi (0.062), film (0.035), stand (0.017), featur (0.015), york (0.015), screen (0.014), live (0.012), comedian (0.012), show (0.012), festiv (0.012), host (0.011), comic (0.01), citi (0.01), laugh (0.008), nyc (0.008), night (0.008), showcas (0.007), ny (0.007), star (0.007), burlesqu (0.007) [$\lambda = 0.3$]: comedi (0.062), film (0.035), comedian (0.012), comic (0.01), burlesqu (0.007), stand (0.017), screen (0.014), filmmak (0.004), laugh (0.008), ny (0.007), york (0.015), standup (0.003), documentari (0.004), show (0.012), funni (0.004), podcast (0.004), cinema (0.004), funniest (0.002), joke (0.003), newfilmmak (0.002)
37	[$\lambda = 1$]: month (0.029), discuss (0.029), club (0.023), share (0.022), coffe (0.02), talk (0.019), topic (0.019), meetup (0.018), speak (0.017), week (0.016), speaker (0.016), skill (0.015), toastmast (0.013), question (0.013), convers (0.011), present (0.011), guest (0.011), peopl (0.01), attend (0.009), support (0.009) [$\lambda = 0.3$]: toastmast (0.013), coffe (0.02), speak (0.017), topic (0.019), discuss (0.029), speech (0.005), speaker (0.016), talk (0.019), club (0.023), month (0.029), wordpress (0.002), communic (0.009), convers (0.011), share (0.022), impromptu (0.001), sps (0.001), evalu (0.003), asl (0.001), skill (0.015), wpn (0.001)
38	[$\lambda = 1$]: startup (0.073), tech (0.055), session (0.034), tool (0.02), market (0.017), process (0.017), system (0.017), explor (0.016), manag (0.016), busi (0.015), workshop (0.014), consult (0.014), capit (0.014), idea (0.013), project (0.013), creativ (0.013), platform (0.012), skill (0.011), strategi (0.011), client (0.011) [$\lambda = 0.3$]: startup (0.073), tech (0.055), hardwar (0.009), capit (0.014), vr (0.007), hack (0.007), prototyp (0.005), 3d (0.006), autom (0.008), platform (0.012), consult (0.014), ar (0.004), tool (0.02), atechup (0.003), fintech (0.003), incub (0.003), holgraph (0.002), session (0.034), analyt (0.006), cybersecur (0.002)
39	[$\lambda = 1$]: park (0.119), street (0.049), locat (0.025), lot (0.023), st (0.018), walk (0.013), build (0.012), south (0.012), ave (0.012), front (0.01), left (0.009), garag (0.009), entranc (0.009), block (0.009), station (0.009), center (0.009), north (0.008), east (0.008), main (0.008), west (0.007)

	[$\lambda = 0.3$]: park (0.119), street (0.049), garag (0.009), ave (0.012), st (0.018), entranc (0.009), lot (0.023), station (0.009), metro (0.005), corner (0.007), block (0.009), south (0.012), meter (0.003), squar (0.007), mardi (0.002), avenu (0.006), lincoln (0.003), locat (0.025), front (0.01), driveway (0.001)
40	[$\lambda = 1$]: perform (0.035), music (0.034), seri (0.024), summer (0.023), live (0.022), theatr (0.018), stage (0.018), camp (0.016), festiv (0.015), austin (0.015), art (0.015), improv (0.013), featur (0.012), theater (0.012), season (0.011), christma (0.011), present (0.01), houston (0.01), product (0.01), of (0.008) [$\lambda = 0.3$]: theatr (0.018), improvis (0.007), camp (0.016), christma (0.011), theater (0.012), austin (0.015), stage (0.018), ballet (0.007), shakespear (0.003), summer (0.023), perform (0.035), seri (0.024), nutcrack (0.002), circus (0.003), carol (0.003), troup (0.002), hisd (0.001), cabaret (0.002), theatric (0.002), balletx (0.001)
41	[$\lambda = 1$]: san (0.087), celebr (0.04), francisco (0.027), diego (0.026), chicago (0.022), year (0.022), antonio (0.018), bay (0.016), cruis (0.015), annual (0.015), parti (0.014), new (0.014), citi (0.012), eve (0.012), at (0.011), anniversari (0.011), hotel (0.01), jose (0.01), in (0.01), yelp (0.009) [$\lambda = 0.3$]: san (0.087), francisco (0.027), diego (0.026), year (0.022), antonio (0.018), cruis (0.015), eve (0.012), jose (0.01), yacht (0.007), bay (0.016), yelp (0.009), anniversari (0.011), celebr (0.04), chicago (0.022), aboard (0.004), sf (0.008), patrick (0.005), parad (0.006), new (0.014), halloween (0.008)
42	[$\lambda = 1$]: adventur (0.04), movi (0.03), hunt (0.021), charact (0.015), scaveng (0.014), watch (0.014), mysteri (0.013), murder (0.012), gaslamp (0.009), hour (0.008), dunnit (0.007), tour (0.007), villag (0.007), world (0.007), video (0.007), quarter (0.007), d (0.006), leagu (0.006), of (0.006), style (0.006) [$\lambda = 0.3$]: adventur (0.04), movi (0.03), hunt (0.021), scaveng (0.014), murder (0.012), mysteri (0.013), charact (0.015), gaslamp (0.009), dunnit (0.007), watson (0.006), d (0.006), crime (0.005), seaport (0.005), spider (0.003), dungeon (0.004), haunt (0.005), dragon (0.006), who (0.005), clue (0.003), 5e (0.002)

43	[$\lambda = 1$]: real (0.056), estat (0.048), invest (0.03), investor (0.02), money (0.016), financi (0.016), deal (0.016), properti (0.011), home (0.011), tax (0.009), incom (0.009), busi (0.008), market (0.008), buy (0.008), educ (0.007), success (0.007), strategi (0.007), month (0.006), flip (0.006), build (0.006) [$\lambda = 0.3$]: estat (0.048), real (0.056), invest (0.03), investor (0.02), financi (0.016), properti (0.011), deal (0.016), incom (0.009), flip (0.006), wealth (0.006), tax (0.009), money (0.016), loan (0.004), mortgag (0.004), debt (0.004), wholesal (0.003), buyer (0.003), retir (0.004), lender (0.002), passiv (0.002)
44	[$\lambda = 1$]: class (0.092), workshop (0.045), student (0.032), session (0.021), skill (0.016), week (0.016), basic (0.016), teach (0.014), program (0.013), hour (0.013), instructor (0.012), hand (0.011), techniqu (0.011), experi (0.011), practic (0.01), level (0.01), train (0.01), provid (0.008), attend (0.008), school (0.007) [$\lambda = 0.3$]: class (0.092), student (0.032), workshop (0.045), basic (0.016), sew (0.003), instructor (0.012), teach (0.014), tuition (0.002), enrol (0.002), teacher (0.007), techniqu (0.011), fundament (0.004), lash (0.001), taught (0.004), skill (0.016), eyelash (0.001), session (0.021), kit (0.003), hand (0.011), microblad (0.001)
45	[$\lambda = 1$]: wine (0.058), beer (0.049), tast (0.028), food (0.027), holiday (0.02), brew (0.02), craft (0.017), festiv (0.017), shop (0.017), market (0.016), local (0.016), enjoy (0.015), breweri (0.014), pop (0.009), cocktail (0.009), vendor (0.009), celebr (0.008), featur (0.008), pair (0.008), at (0.008) [$\lambda = 0.3$]: wine (0.058), beer (0.049), tast (0.028), brew (0.02), breweri (0.014), wineri (0.007), craft (0.017), holiday (0.02), truck (0.007), shop (0.017), whiskey (0.005), cider (0.003), vendor (0.009), fest (0.007), pearl (0.005), farmer (0.005), sampl (0.007), ros (0.002), warehous (0.004), pour (0.004)
46	[$\lambda = 1$]: mask (0.037), inform (0.021), pm (0.018), job (0.017), www (0.017), call (0.017), contact (0.016), lunch (0.016), fair (0.016), servic (0.015), visit (0.013), locat (0.013), career (0.012), attend (0.01), chapter (0.009), opportun (0.008), email (0.008), hire (0.007), center (0.007), org (0.007)

	[$\lambda = 0.3$]: fair (0.016), mask (0.037), job (0.017), resum (0.004), hire (0.007), employ (0.007), lunch (0.016), interview (0.007), pa (0.005), az (0.003), www (0.017), phoenix (0.006), seeker (0.002), chapter (0.009), arizona (0.003), wexford (0.001), career (0.012), scottsdal (0.002), wsba (0.001), tarrant (0.001)
47	[$\lambda = 1$]: and (0.044), a (0.041), to (0.033), dog (0.027), with (0.024), for (0.021), of (0.019), in (0.015), present (0.013), on (0.011), host (0.011), you (0.01), your (0.01), pet (0.01), cat (0.009), show (0.009), anim (0.009), it (0.008), is (0.007), up (0.007) [$\lambda = 0.3$]: dog (0.027), a (0.041), cat (0.009), you (0.01), pet (0.01), to (0.033), m (0.005), is (0.007), i (0.007), it (0.008), puppi (0.003), and (0.044), farrel (0.003), h (0.003), pup (0.003), edward (0.005), are (0.004), rescu (0.005), paw (0.002), what (0.004)
48	[$\lambda = 1$]: ticket (0.031), email (0.018), question (0.017), refund (0.017), registr (0.016), contact (0.015), requir (0.015), purchas (0.013), fee (0.012), inform (0.012), cancel (0.011), hour (0.011), regist (0.01), includ (0.01), date (0.01), option (0.009), check (0.009), attend (0.009), person (0.009), reserv (0.009) [$\lambda = 0.3$]: refund (0.017), ticket (0.031), email (0.018), faq (0.006), registr (0.016), id (0.007), transfer (0.003), polici (0.008), fee (0.012), cancel (0.011), shoot (0.004), prior (0.008), purchas (0.013), print (0.007), payment (0.006), eventbrit (0.005), minimum (0.005), transferr (0.001), paypal (0.002), firearm (0.001)

Table C.1: Topics and their word distributions generated from LDA against 179,858 random samples of the event datasets. The top-20 words for each topic are presented in the descending order of proportional weight. λ is the relevance score where $\lambda = 1$ is the original word distribution of a topic, and $\lambda = 0.3$ presents a word distribution that is specific to the corresponding topic (so represents the topic better). See [Sievert and Shirley, 2014] for details.

Appendix D: Cultural Activity Diversity of Target Cities over Time

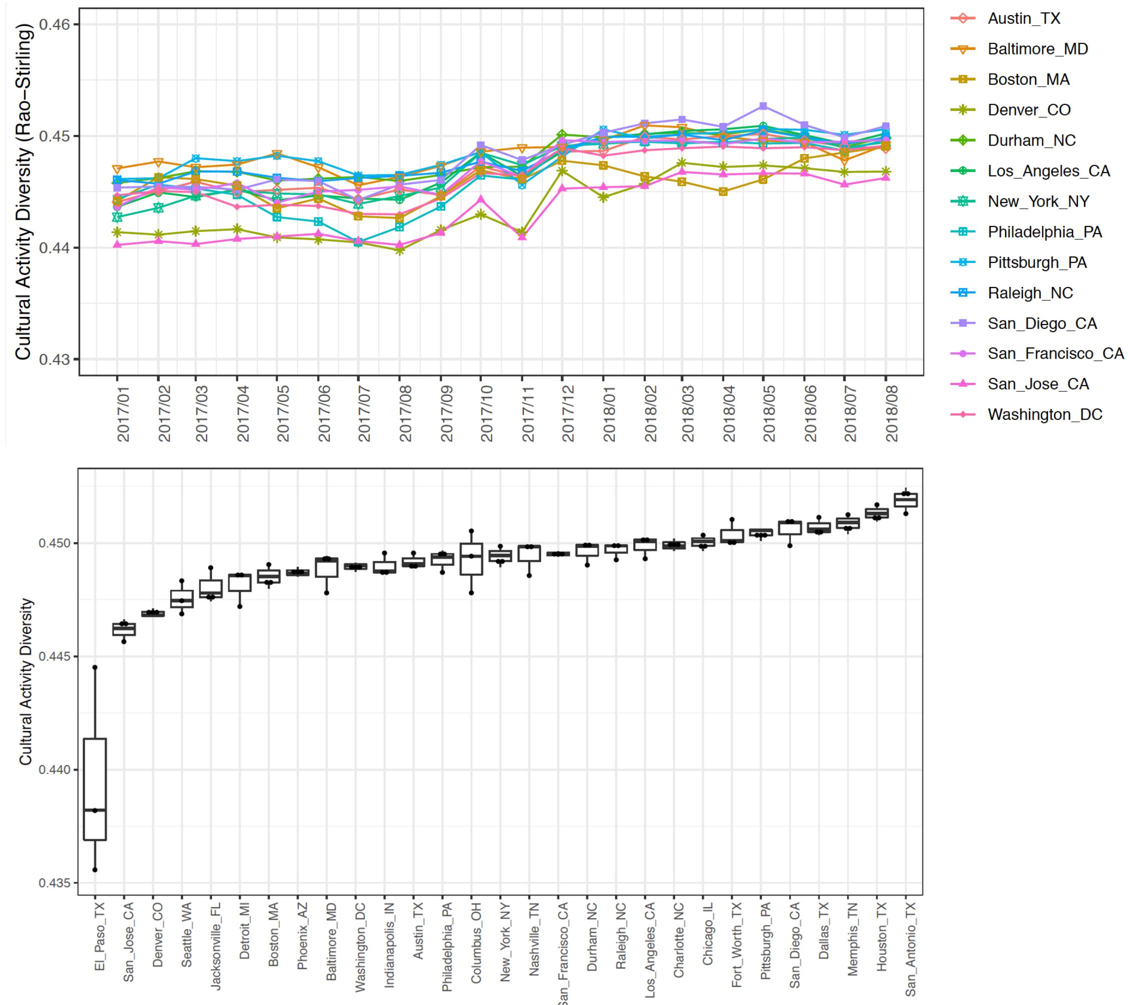

Figure D.1: Cultural activity diversity changes (Rao-Stirling) in 14 cities over 20 months and 29 cities over 3 months.

Appendix E: Unweighted Descriptive Statistics for Variables

	Mean	SD
Ethnoracial heterogeneity (HHI)	.586	.032
Socio-economic segregation (Gini)	.460	0.032
Socio-economic deprivation		
Rate below poverty level	.146	.048
Rate rented households	.395	.071
Rate unemployment below poverty level	.203	.070
Rate unemployment at/above poverty level	.036	.013
Rate education less than bachelor's	.666	.104
Deprivation index	.608	.111
Socio-demographic features		
Rate adults over 65y/o	.209	.047
Median age	36.6	4.78
Rate female over 16y/o	.515	.014
Control variables		
Rate white	.733	.128
Rate black/African-American	.126	.112
Rate Native American/Alaskan	.006	.009
Rate Asian	.050	.057
Rate Hawaiian/Pacific islanders	.003	.009
Rate others	.043	.054
Rate foreign-born	.007	.005

Rate U.S. citizens	. . . .	.939	.039
City population	. . . .	802814	1750606
Commuting time (min)	. . . .	23.8	5.07
Median income	. . . .	59251	14221

Table E.1: Descriptive statistics for the census-based variables of 269 urban areas in the U.S. in 2017.

		Mean	SD
Total events	. . . .	6683	6119
The fragmentation of event information (β_{CO})	. . . .	.932	.036
Cultural activity diversity	. . . .		
Rao-Stirling index	. . . .	.447	.003
Rao-Stirling index ($\beta = 0.5$)	. . . .	19.3	.556
Rao-Stirling index ($\alpha = 0.5$)	. . . .	.465	.002
Shannon entropy	. . . .	3.65	.052
HHI	. . . .	.031	.003
Community engagement	. . . .		
Standardized number of events /population	. . . .	.006	.004
Standardized number of RSVPs /event	. . . .	12.5	2.82
Standardized number of RSVPs /population	. . . .	.045	.029

Table E.2: Descriptive statistics for the event data-driven variables for 14 cities over 20 months.

	Mean	SD
Total events	7024	7164
The fragmentation of event information (β_{CO})	.907	.032
Cultural activity diversity		
Rao-Stirling index	.449	.002
Rao-Stirling index ($\beta = 0.5$)	19.82	.503
Rao-Stirling index ($\alpha = 0.5$)	.467	.002
Shannon entropy	3.70	.055
HHI	.029	.005
Community engagement		
Standardized number of events /population	.006	.005
Standardized number of RSVPs /event	11.0	2.39
Standardized number of RSVPs /population	.027	.025

Table E.3: Descriptive statistics for the event data-driven variables for 29 cities over 3 months.

Appendix F: Effect Size Changes over Time

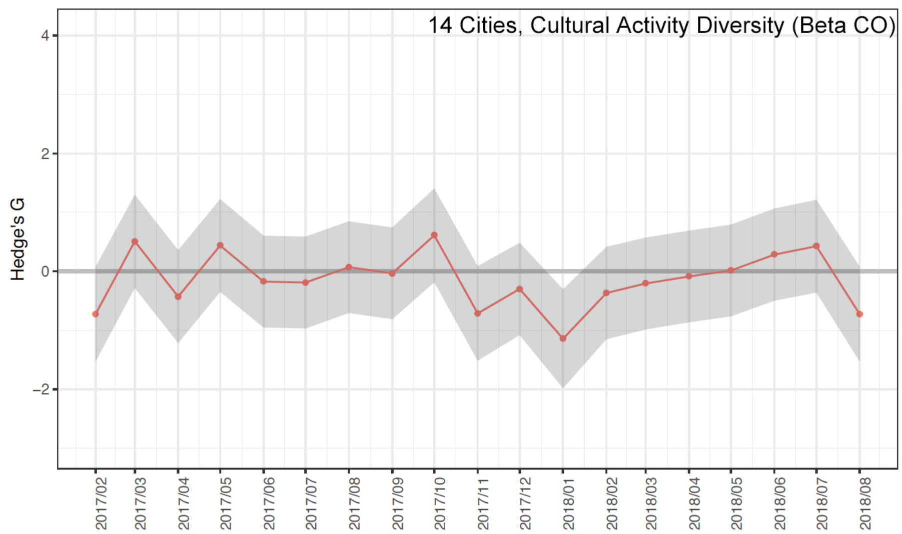

Figure F.1: Hedge's D for the fragmentation of local information (β_{CO}) in 14 cities over 20 months. Each month's value is Hedge's D compared to the previous month's distribution. The gray area shows confidence intervals. Except for January 2018, there is no systematic change.

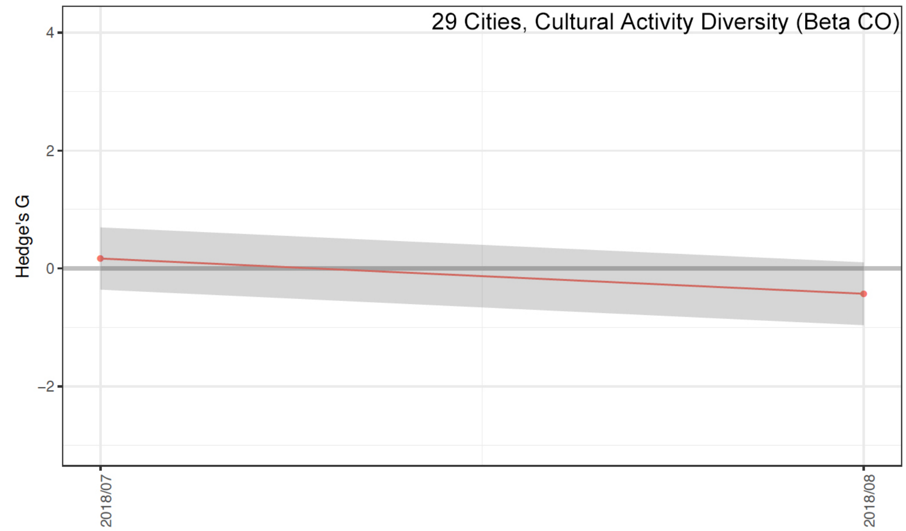

Figure F.2: Hedge's D for the fragmentation of local information (β_{CO}) in 29 cities over 3 months. Each month's value is Hedge's D compared to the previous month's distribution. The gray area shows confidence intervals. There is no systematic change over time.

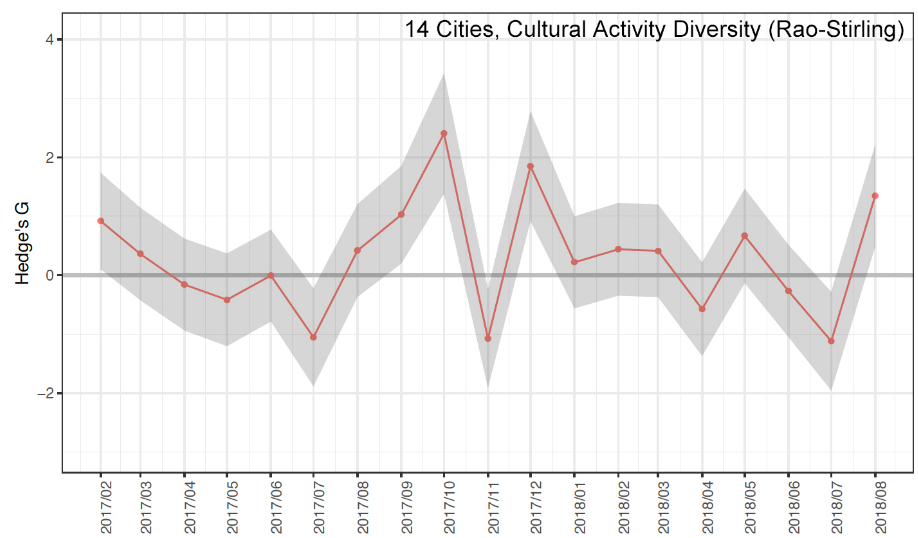

Figure F.3: Hedge's D for 14 cities' cultural activity diversity scores (Rao-Stirling index) over 20 months. Each month's value is Hedge's D compared to the previous month's distribution. The gray area shows confidence intervals. There are some fluctuations in the fall, but not a linear change over time.

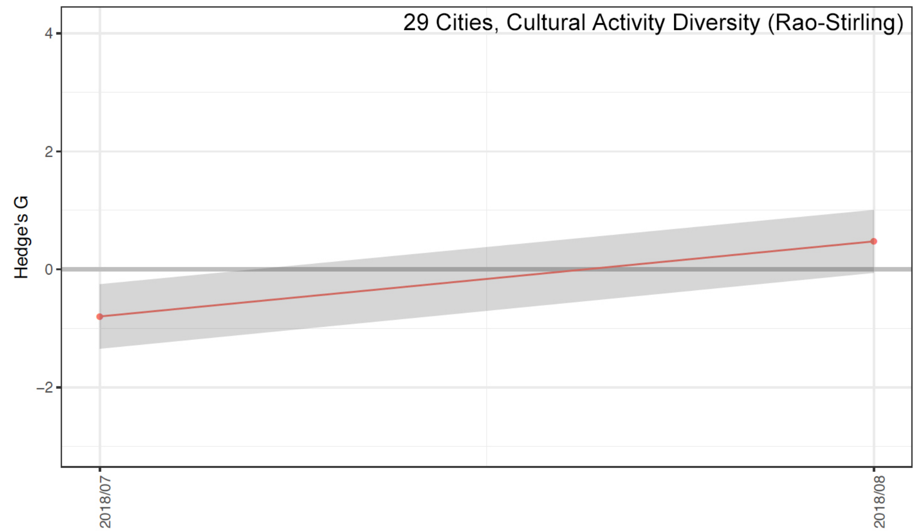

Figure F.4: Hedge's D for 29 cities' cultural activity diversity scores (Rao-Stirling index) over 3 months. Each month's value is Hedge's D compared to the previous month's distribution. The gray area shows confidence intervals. There is a decrease in July .

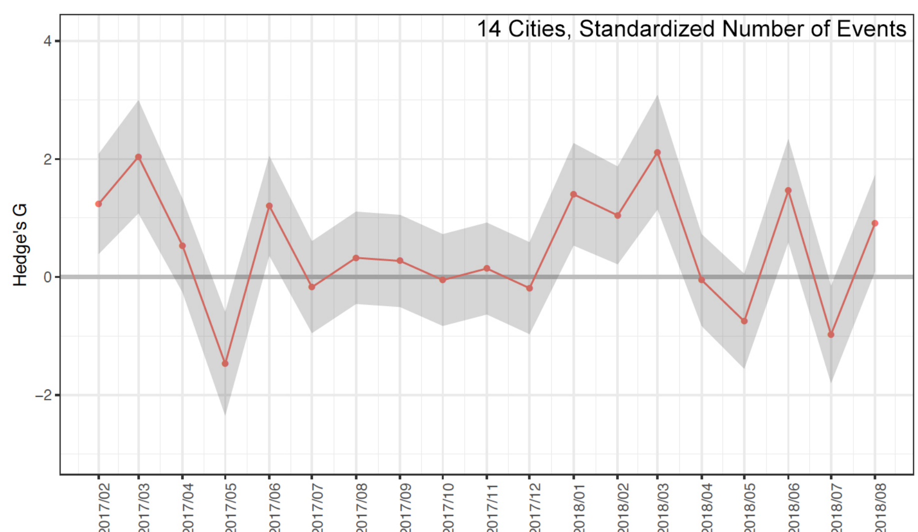

Figure F.5: Hedge's D for 14 cities' numbers of events normalized by population over 20 months. Each month's value is Hedge's D compared to the previous month's distribution. The gray area shows confidence intervals. There are seasonal effects (more events in Spring).

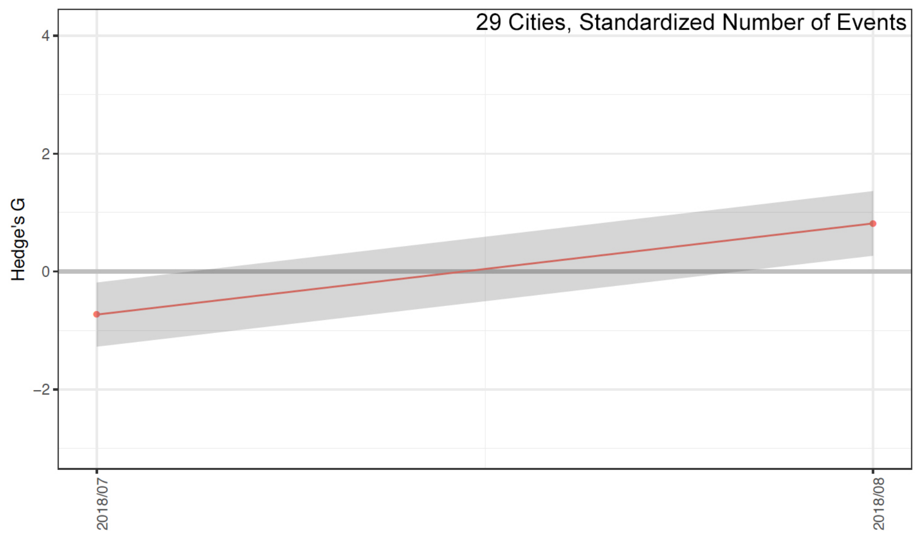

Figure F.6: Hedge's D for 29 cities' numbers of events normalized by population over 3 months. Each month's value is Hedge's D compared to the previous month's distribution. The gray area shows confidence intervals. There are meaningful changes but inconsistent (decreases in July and increases in August).

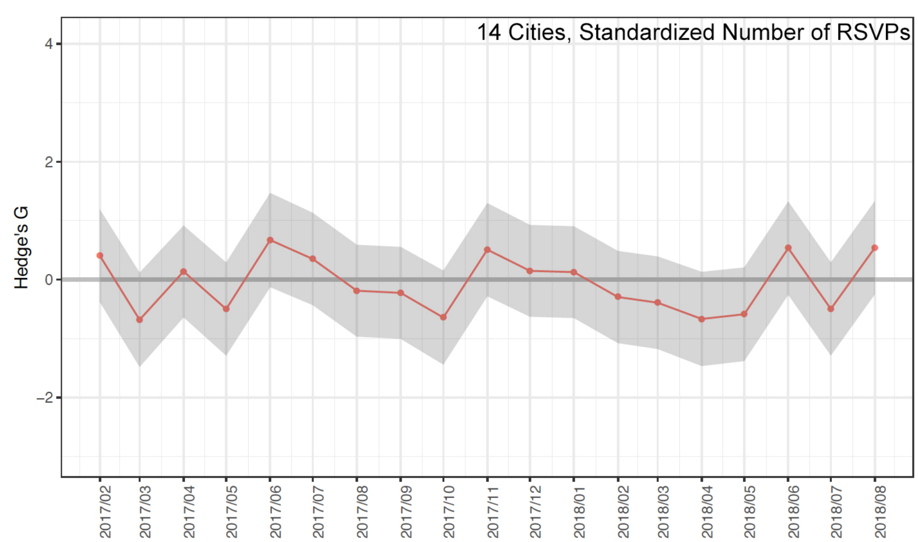

Figure F.7: Hedge's D for 14 cities' numbers of RSVPs normalized by events over 20 months. Each month's value is Hedge's D compared to the previous month's distribution. The gray area shows confidence intervals. There is no systematic change over time.

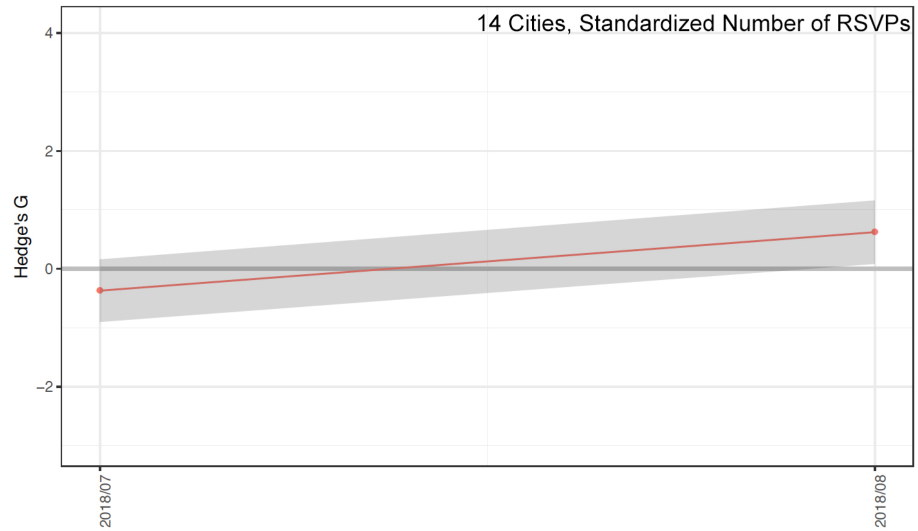

Figure F.8: Hedge's D for 29 cities' numbers of RSVPs normalized by events over 3 months. Each month's value is Hedge's D compared to the previous month's distribution. The gray area shows confidence intervals. There is an increase in the value in August.

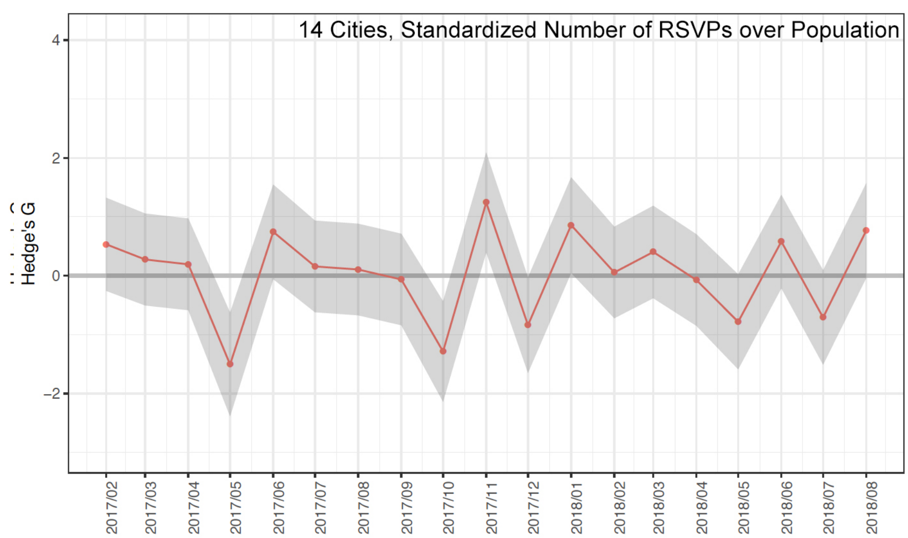

Figure F.9: Hedge's D for 14 cities' numbers of RSVPs normalized by population over 20 months. Each month's value is Hedge's D compared to the previous month's distribution. The gray area shows confidence intervals. There are some changes in between, but random.

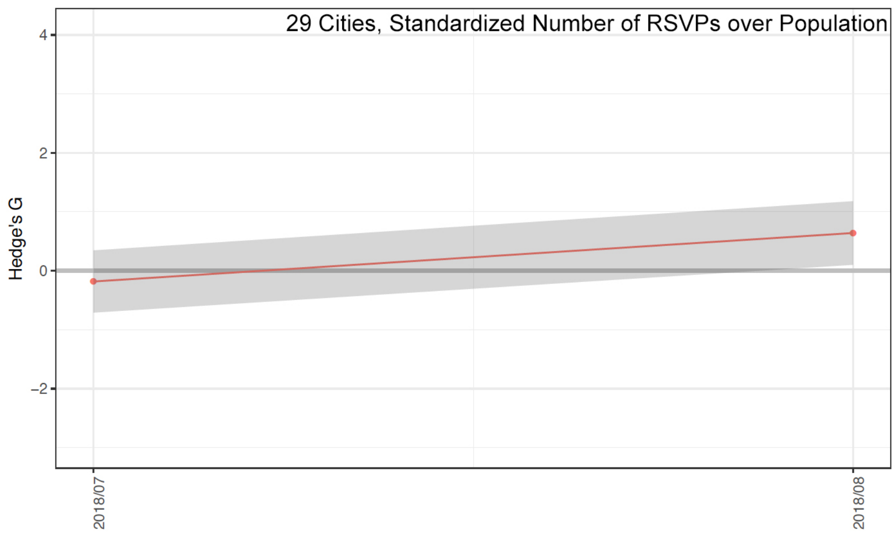

Figure F.10: Hedge's D for 29 cities' numbers of RSVPs normalized by population over 3 months. Each month's value is Hedge's D compared to the previous month's distribution. The gray area shows confidence intervals. There is an increase in August.

Appendix G: Regression Results

The followings are the complete list of regression tables that are discussed in Chapter 5. All the regressions starting from H1-2 results make use of Bayesian approach based on MCMC for estimating the distribution of parameters. Analytical models that Bayesian-based regressions test are same as the regular linear multilevel models (LMM), but only the parameter estimation method is different. This approach is driven by the small number of data points within the random-effect variable (if using regular LMM models, there are often overfitting and multicollinearity issues present).

The standard confidence interval (CI) in the results of Bayesian-based LMMs is 90% confidence, instead of 95%. This is because of the characteristic of estimation methods. While CI in regular LMMs assures the confidence of the estimation conditional to the given data and model, the meaning of CI in Bayesian-based LMMs is that the estimate of a value γ is within its 100*p% posterior interval [Gabry and Goodrich, 2018]. With this different meaning of CI, scholars suggest 90% as the reliable criteria instead of 95% for ensuring the computational stability and sign errors [Gelman and Carlin, 2014]. Therefore, the regression tables report the estimates of parameters, 5% CI, and 95% CI throughout the report.

	Model Gini-1			Model Depr-1			Model Gini-2			Model Depr-2		
	Est.	S.E.	p	Est.	S.E.	p	Est.	S.E.	p	Est.	S.E.	p
H1-1 (SES): Census Estimate ($N = 269$)												
(Intercept)	0.227	0.084	0.007***	0.261	0.284	0.360	0.479	0.195	0.015**	-0.056	0.654	0.931
Ethnoracial heterogeneity (HHI)	0.227	0.065	0.001***	0.622	0.222	0.006***	0.178	0.158	0.260	0.256	0.529	0.629
Rate foreign-born	0.532	0.344	0.124	-9.895	1.222	0.000***	-0.040	0.359	0.911	-6.340	1.200	0.000***
Rate U.S. citizens	-0.111	0.054	0.042**	-0.047	0.182	0.798	-0.228	0.069	0.001***	0.122	0.230	0.597
Rate female over 16	0.551	0.127	0.000***	-1.228	0.460	0.008***	0.320	0.130	0.015**	-0.960	0.440	0.030**
Rate adults over 65	0.381	0.102	0.000***	0.954	0.357	0.008***	0.418	0.098	0.000***	0.812	0.325	0.013**
Median age	-0.003	0.001	0.002***	-0.004	0.004	0.277	-0.003	0.001	0.001***	-0.003	0.003	0.427
Commuting time (min)	-0.002	0.000	0.000***	0.003	0.002	0.034**	-0.002	0.000	0.000***	0.004	0.001	0.005***
City population	0.000	0.000	0.000***	-0.000	0.000	0.009***	0.000	0.000	0.001***	-0.000	0.000	0.040**
Rate white							0.006	0.123	0.959	0.154	0.413	0.709
Rate black/African-American							0.044	0.122	0.722	0.304	0.409	0.457
Rate Native American/Alaskan							-0.196	0.256	0.445	0.941	0.852	0.270
Rate Asian							0.104	0.142	0.465	-0.729	0.479	0.129
Rate Hawaiian/Pacific islanders							-0.706	0.418	0.093**	3.323	1.386	0.017**
Rate others							-0.129	0.135	0.340	0.752	0.452	0.097**

Table G.1: Regression results for testing H1-1 about the correlation between socio-economic deprivation/inequality and ethnoracial heterogeneity. The dependent variables, Gini index and deprivation index, are nested within the state.

	Model HHI-1			Model HHI-2			Model HHI-3			Model HHI-4		
	Est.	S.E.	p	Est.	S.E.	p	Est.	S.E.	p	Est.	S.E.	p
H1-1 (HHI): Census Estimate ($N = 269$)												
(Intercept)	0.398	0.073	0.000***	0.338	0.077	0.000***	0.637	0.067	0.000***	0.599	0.069	0.000***
Socio-economic inequality (Gini)	0.172	0.054	0.002***	0.180	0.067	0.008***	0.031	0.025	0.223	0.066	0.033	0.049**
Socio-economic deprivation (depr. index)	0.045	0.014	0.002***				0.003	0.008	0.664			
Rate foreign-born	0.748	0.319	0.020**	0.324	0.318	0.310	0.148	0.150	0.324	0.167	0.156	0.288
Rate U.S. citizens	-0.052	0.047	0.271	-0.091	0.044	0.040**	-0.014	0.028	0.616	-0.013	0.028	0.643
Rate female over 16	0.297	0.107	0.006***	0.351	0.101	0.001***	0.001	0.052	0.979	0.026	0.052	0.619
Median age	0.002	0.001	0.073**	0.001	0.001	0.144	0.001	0.000	0.146	0.000	0.000	0.845
Rate adults over 65	-0.400	0.087	0.000***	-0.347	0.086	0.000***	-0.087	0.041	0.034**	-0.067	0.043	0.124
Commuting time (min)	0.001	0.000	0.024**	0.001	0.000	0.080**	0.000	0.000	0.098**	0.000	0.000	0.422
City population	0.000	0.000	0.000***	0.000	0.000	0.007***	0.000	0.000	0.011**	0.000	0.000	0.024**
Rate below poverty level				-0.146	0.053	0.006***				-0.059	0.026	0.026**
Rate rented households				0.113	0.030	0.000***				0.003	0.014	0.846
Rate unemployment below poverty level				0.053	0.021	0.014**				0.008	0.011	0.468
Rate unemployment at/above poverty level				0.516	0.129	0.000***				0.158	0.065	0.016**
Rate education less than bachelor's				0.021	0.019	0.280				0.005	0.010	0.662
Rate white							-0.122	0.048	0.012**	-0.099	0.049	0.045**
Rate black/African-American							0.108	0.048	0.027**	0.124	0.048	0.011**
Rate Native American/Alaskan							0.332	0.101	0.001***	0.339	0.100	0.001***
Rate Asian							0.110	0.056	0.052**	0.133	0.056	0.019**
Rate Hawaiian/Pacific islanders							-0.538	0.166	0.001***	-0.503	0.169	0.003***
Rate others							0.104	0.053	0.052**	0.124	0.053	0.021**

Table G.2: Regression results for testing H1-1 about the correlation between socio-economic deprivation/inequality and ethno-racial heterogeneity. The dependent variable, HHI, is nested within the state.

	Std. Events/Pop			Std. RSVPs/Pop			Std. RSVPs/Event		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
Comm. Engagement Baseline: 14 Areas ($N = 280$)									
(Intercept)	-0.004	-0.209	0.187	0.002	-0.083	0.085	0.001	-0.059	0.062
Rate foreign-born	1.828	1.550	2.107	2.927	2.744	3.098	0.802	0.600	0.993
Rate U.S. citizens	0.978	0.411	1.542	2.598	2.244	2.925	1.136	0.732	1.525
Rate female over	0.274	0.101	0.448	0.278	0.179	0.383	0.458	0.343	0.574
Median age	2.149	1.868	2.435	3.101	2.943	3.255	1.062	0.874	1.245
Rate adults over 65	-1.594	-1.874	-1.305	-2.707	-2.861	-2.543	-1.371	-1.556	-1.179
Commuting time (min)	0.247	0.130	0.361	0.162	0.101	0.226	0.181	0.105	0.260
Rate white	8.873	6.847	10.996	17.436	15.859	18.888	0.261	-1.235	1.751
Rate black/African-American	7.381	5.733	9.118	14.236	12.956	15.412	-0.070	-1.305	1.139
Rate Native American/Alaskan	0.964	0.834	1.098	1.474	1.397	1.548	0.285	0.201	0.370
Rate Asian	5.621	4.141	7.191	12.451	11.284	13.541	1.117	-0.015	2.233
Rate Hawaiian/Pacific islanders	1.103	0.921	1.288	2.005	1.871	2.131	0.261	0.127	0.388
Rate others	4.117	3.219	5.049	7.890	7.191	8.527	-0.116	-0.785	0.552
City population	-0.226	-0.401	-0.050	0.271	0.170	0.369	0.989	0.868	1.104

Table G.3: Baseline regression results for testing the effects of control variables on community engagement indicators for 14 areas over 20 months. All the variables are scaled to see the relative effects. Community engagement indicators are nested within the time (random effects).

	Std. Events/Pop			Std. RSVPs/Pop			Std. RSVPs/Event		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
Comm. Engagement Baseline: 28 Areas ($N = 84$)									
(Intercept)	0.023	-0.170	0.206	0.024	-0.162	0.209	0.005	-0.151	0.156
Rate foreign-born	0.670	0.439	0.898	0.655	0.428	0.892	0.002	-0.147	0.153
Rate U.S. citizens	0.147	-0.255	0.560	0.219	-0.188	0.623	0.099	-0.167	0.362
Rate female over	0.222	-0.024	0.462	0.357	0.106	0.611	0.174	0.027	0.320
Median age	0.636	0.264	1.006	0.636	0.253	1.002	0.082	-0.127	0.290
Rate adults over 65	-0.419	-0.740	-0.089	-0.570	-0.897	-0.243	-0.338	-0.530	-0.146
Commuting time (min)	0.611	0.371	0.855	0.507	0.251	0.756	0.314	0.172	0.453
Rate white	-0.878	-3.090	1.401	-0.162	-2.502	2.091	-0.153	-2.078	1.820
Rate black/African-American	-0.759	-2.932	1.430	-0.174	-2.418	1.966	-0.374	-2.223	1.531
Rate Native American/Alaskan	0.130	-0.040	0.296	0.221	0.051	0.395	-0.002	-0.117	0.112
Rate Asian	-0.967	-2.580	0.680	-0.186	-1.871	1.445	0.699	-0.700	2.133
Rate Hawaiian/Pacific islanders	0.115	-0.123	0.353	0.115	-0.141	0.357	-0.112	-0.290	0.068
Rate others	-0.131	-1.110	0.863	0.158	-0.844	1.138	-0.435	-1.269	0.417
City population	-0.496	-0.698	-0.297	-0.453	-0.660	-0.245	0.480	0.359	0.598

Table G.4: Baseline regression results for testing the effects of control variables on community engagement indicators for for 28 areas over 3 months. All the variables are scaled to see the relative effects. Community engagement indicators are nested within the time (random effects).

	Model		
	Est.	CI 5%	CI 95%
Frag. Baseline: 14 Areas ($N = 280$)			
(Intercept)	0.005	-0.330	0.342
Rate foreign-born	0.159	-0.171	0.496
Rate U.S. citizens	0.031	-0.667	0.743
Rate female over 16	0.336	0.129	0.534
Median age	0.873	0.517	1.240
Rate adults over 65	-0.717	-1.084	-0.363
Commuting time (min)	-0.519	-0.667	-0.373
Rate white	1.399	-0.648	3.435
Rate black/African-American	1.140	-0.530	2.773
Rate Native American/Alaskan	0.361	0.204	0.519
Rate Asian	1.394	-0.134	2.885
Rate Hawaiian/Pacific islanders	0.167	-0.019	0.354
Rate others	0.052	-0.851	0.943
City population	0.448	0.230	0.674
sigma	0.463	0.430	0.498
mean_PPD	0.001		
log-posterior	-229.761		

Table G.5: Baseline regression results for testing the effects of control variables on the fragmentation of local information for for 14 areas over 20 months. All the variables are scaled to see the relative effects. Fragmentation indicators are nested within the time (random effects).

	Model		
	Est.	CI 5%	CI 95%
Frag. Baseline: 28 Areas ($N = 84$)			
(Intercept)	-0.025	-0.307	0.283
Rate foreign-born	-0.001	-0.285	0.272
Rate U.S. citizens	0.197	-0.303	0.680
Rate female over 16	0.154	-0.162	0.477
Median age	-0.700	-1.167	-0.243
Rate adults over 65	0.652	0.240	1.065
Commuting time (min)	-0.328	-0.636	-0.018
Rate white	0.799	-1.617	3.147
Rate black/African-American	0.586	-1.708	2.850
Rate Native American/Alaskan	0.379	0.176	0.577
Rate Asian	0.965	-0.762	2.662
Rate Hawaiian/Pacific islanders	-0.040	-0.317	0.230
Rate others	0.289	-0.759	1.327
City population	-0.151	-0.415	0.111
sigma	0.764	0.666	0.878
mean_PPD	-0.032		
log-posterior	-120.835		

Table G.6: Baseline regression results for testing the effects of control variables on the fragmentation of local information for for 28 areas over 3 months. All the variables are scaled to see the relative effects. Fragmentation indicators are nested within the time (random effects).

	Rao-Stirling			RS β-weighted			RS α-weighted			Shannon Entropy			HHI		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
Cultural/Base 14 Areas ($N = 280$)															
(Intercept)	-0.007	-0.309	0.288	-0.007	-0.315	0.296	0.009	-0.320	0.331	0.004	-0.299	0.313	-0.011	-0.317	0.301
Rate foreign-born	-0.892	-1.162	-0.622	-0.927	-1.211	-0.630	-0.313	-0.589	-0.040	-0.609	-0.900	-0.315	-0.909	-1.235	-0.570
Rate U.S. citizens	-1.593	-2.138	-1.057	-1.812	-2.392	-1.199	-0.698	-1.267	-0.143	-1.394	-2.006	-0.812	-2.030	-2.726	-1.320
Rate female over 16	-0.201	-0.362	-0.038	-0.184	-0.362	-0.009	-0.473	-0.636	-0.311	-0.355	-0.530	-0.180	-0.152	-0.360	0.042
Median age	-0.173	-0.433	0.083	-0.174	-0.459	0.106	-0.155	-0.410	0.102	-0.192	-0.490	0.104	-0.206	-0.553	0.151
Rate adults over 65	0.429	0.167	0.691	0.518	0.230	0.814	0.486	0.223	0.749	0.625	0.328	0.925	0.674	0.337	1.019
Commuting time (min)	-0.452	-0.558	-0.345	-0.403	-0.519	-0.284	-0.252	-0.358	-0.145	-0.253	-0.368	-0.136	-0.243	-0.388	-0.099
Rate white	-9.099	-11.061	-7.145	-7.602	-9.631	-5.504	-5.270	-7.294	-3.360	-4.800	-6.791	-2.891	-4.301	-6.384	-2.159
Rate black/African-American	-6.760	-8.372	-5.172	-5.571	-7.218	-3.852	-3.644	-5.287	-2.087	-3.292	-4.882	-1.728	-2.987	-4.688	-1.248
Rate Native American/Alaskan	-0.795	-0.912	-0.676	-0.696	-0.825	-0.565	-0.534	-0.654	-0.416	-0.507	-0.637	-0.374	-0.486	-0.641	-0.333
Rate Asian	-7.699	-9.189	-6.228	-6.546	-8.043	-5.002	-4.445	-5.961	-3.015	-4.231	-5.737	-2.800	-3.941	-5.524	-2.313
Rate Hawaiian/Pacific islanders	-0.194	-0.369	-0.020	-0.167	-0.345	0.024	0.005	-0.173	0.178	-0.040	-0.218	0.133	-0.107	-0.299	0.093
Rate others	-3.865	-4.774	-2.990	-3.337	-4.239	-2.413	-2.001	-2.906	-1.144	-2.053	-2.958	-1.169	-2.078	-3.034	-1.108
City population	-0.070	-0.237	0.091	-0.166	-0.346	0.016	0.043	-0.127	0.210	-0.155	-0.339	0.028	-0.354	-0.566	-0.139

Table G.7: Baseline regression results for testing the effects of control variables on cultural activity diversity for for 14 areas over 20 months. All the variables are scaled to see the relative effects. Diversity indicators are nested within the time (random effects).

	Rao-Stirling			RS β-weighted			RS α-weighted			Shannon Entropy			HHI		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
Cultural (base): 28 Areas ($N = 84$)															
(Intercept)	-0.016	-0.365	0.339	-0.023	-0.340	0.297	-0.029	-0.394	0.312	-0.012	-0.309	0.303	-0.022	-0.367	0.284
Rate foreign-born	0.031	-0.279	0.352	0.149	-0.150	0.449	0.117	-0.195	0.437	0.178	-0.111	0.465	0.263	-0.009	0.537
Rate U.S. citizens	0.071	-0.488	0.646	0.157	-0.392	0.691	-0.143	-0.692	0.408	-0.007	-0.529	0.509	0.243	-0.236	0.713
Rate female over 16	-0.137	-0.512	0.240	-0.198	-0.536	0.130	-0.426	-0.778	-0.073	-0.359	-0.687	-0.028	-0.253	-0.566	0.064
Median age	0.340	-0.193	0.858	0.450	-0.033	0.929	0.333	-0.175	0.841	0.430	-0.047	0.917	0.564	0.124	1.012
Rate adults over 65	-0.446	-0.915	0.026	-0.484	-0.898	-0.052	-0.210	-0.661	0.237	-0.322	-0.733	0.089	-0.517	-0.915	-0.125
Commuting time (min)	-0.280	-0.621	0.059	-0.297	-0.613	0.027	-0.355	-0.688	-0.011	-0.331	-0.654	-0.013	-0.294	-0.591	0.002
Rate white	-0.472	-2.893	1.907	-0.395	-2.862	2.014	-0.249	-2.662	2.210	-0.192	-2.577	2.252	-0.295	-2.523	2.092
Rate black/African-American	-0.195	-2.491	2.113	-0.017	-2.416	2.328	0.354	-1.967	2.698	0.382	-1.908	2.681	0.170	-2.010	2.456
Rate Native American/Alaskan	-0.024	-0.254	0.206	0.099	-0.120	0.315	0.149	-0.071	0.378	0.203	-0.015	0.419	0.228	0.030	0.425
Rate Asian	-0.223	-2.014	1.529	-0.064	-1.888	1.682	-0.017	-1.791	1.775	0.096	-1.674	1.882	0.118	-1.522	1.872
Rate Hawaiian/Pacific islanders	0.071	-0.224	0.372	0.004	-0.286	0.297	-0.031	-0.328	0.260	-0.055	-0.330	0.222	-0.072	-0.342	0.204
Rate others	-0.730	-1.794	0.364	-0.701	-1.783	0.374	-0.596	-1.674	0.467	-0.588	-1.621	0.462	-0.638	-1.627	0.393
City population	0.678	0.374	0.979	0.705	0.432	0.978	0.595	0.301	0.881	0.624	0.356	0.899	0.690	0.446	0.944

Table G.8: Baseline regression results for testing the effects of control variables on cultural activity diversity for for 28 areas over 3 months. All the variables are scaled to see the relative effects. Diversity indicators are nested within the time (random effects).

	Std. Events/Pop			Std. RSVPs/Pop			Std. RSVPs/Event		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H1-2 (SES): 14 Areas ($N = 280$)									
(Intercept)	-127.182	-180.165	-75.134	-175.481	-216.593	-133.967	-45.529	-91.842	0.802
Socio-economic inequality (Gini)	-38.867	-56.031	-22.054	-47.693	-61.283	-34.104	-6.137	-21.173	8.605
Socio-economic deprivation (depr. index)	-32.762	-37.973	-27.548	-48.705	-52.954	-44.473	-8.212	-12.841	-3.376
Rate foreign-born	404.070	241.574	568.175	496.995	364.219	628.260	146.904	1.382	294.431
Rate U.S. citizens	163.598	121.551	204.860	235.581	204.303	266.330	51.228	16.110	86.931
Rate female over 16	15.451	-53.682	85.681	1.591	-57.778	61.365	22.254	-41.757	87.312
Median age	-2.015	-2.297	-1.737	-2.945	-3.134	-2.746	-0.230	-0.459	-0.005
Rate adults over 65	121.970	91.610	152.573	172.522	146.920	198.287	-1.466	-30.322	25.844
Commuting time (min)	0.444	0.383	0.506	0.559	0.519	0.599	0.133	0.085	0.182
Rate white	16.800	0.768	32.683	28.125	14.009	42.027	-9.156	-24.412	5.828
Rate black/African-American	9.528	-11.915	30.515	18.592	-0.139	37.319	-13.585	-33.819	6.242
Rate Asian	35.096	23.712	46.528	60.116	51.217	68.807	5.511	-4.447	15.048
Rate others	91.464	75.830	107.039	130.497	118.767	141.781	0.013	-13.109	12.653
City population	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table G.9: Regression results for H1-2 regarding community engagement for 14 areas over 20 months. to see whether socio-economic deprivation/inequality indices are correlated to community engagement indicators. DVs are nested within the time.

	Std. Events/Pop			Std. RSVPs/Pop			Std. RSVPs/Event		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H1-2 (SES): 28 Areas ($N = 84$)									
(Intercept)	25.620	10.302	40.576	23.385	7.247	38.980	-10.552	-21.637	0.377
Socio-economic inequality (Gini)	-8.318	-16.198	-0.646	-4.144	-11.869	3.879	9.053	3.362	14.788
Socio-economic deprivation (depr. index)	-7.560	-9.845	-5.226	-8.687	-11.004	-6.334	-2.883	-4.530	-1.275
Rate foreign-born	86.616	40.389	132.439	68.764	21.259	117.880	-10.532	-44.952	25.102
Rate U.S. citizens	-5.782	-12.952	1.732	-5.818	-13.169	1.715	3.140	-1.941	8.456
Rate female over 16	-15.566	-40.409	8.926	-16.207	-41.707	8.915	-8.418	-25.836	9.116
Median age	-0.084	-0.256	0.094	-0.117	-0.288	0.062	-0.037	-0.158	0.083
Rate adults over 65	25.921	12.359	39.793	23.998	9.778	37.450	-6.439	-16.114	3.096
Commuting time (min)	0.122	0.067	0.176	0.086	0.030	0.142	0.089	0.051	0.128
Rate white	-14.973	-24.034	-5.686	-11.594	-20.836	-2.117	7.693	0.701	14.646
Rate black/African-American	-11.282	-20.507	-1.815	-8.423	-17.834	1.016	6.544	-0.498	13.637
Rate Asian	-23.605	-34.125	-13.056	-18.113	-28.706	-7.487	15.829	8.003	23.675
Rate others	-8.268	-18.614	2.269	-4.652	-15.021	5.699	1.503	-6.258	9.135
City population	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000

Table G.10: Regression results for testing H1-2 in 28 areas over 3 months to see whether socio-economic deprivation/inequality indices are correlated to community engagement indicators. DVs are nested within the time.

	Std. Events/Pop			Std. RSVPs/Pop			Std. RSVPs/Event		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H1-2 (SES): 14 Areas ($N = 42$)									
(Intercept)	-151.200	-201.409	-99.070	-141.361	-204.500	-74.403	1.736	-63.051	68.800
Socio-economic inequality (Gini)	-47.571	-63.836	-30.973	-39.590	-59.101	-19.389	7.153	-13.347	27.731
Socio-economic deprivation (depr. index)	-40.693	-46.548	-34.675	-43.784	-50.490	-36.340	-10.841	-18.194	-3.495
Rate foreign-born	531.011	367.163	689.349	414.365	217.066	608.569	-11.551	-213.278	186.265
Rate U.S. citizens	205.575	163.797	245.426	196.972	138.564	249.594	16.139	-38.959	69.326
Rate female over 16	-6.224	-78.135	64.642	4.001	-71.760	82.742	-27.445	-108.597	53.450
Median age	-2.489	-2.793	-2.137	-2.472	-2.890	-1.976	-0.349	-0.787	0.103
Rate adults over 65	155.985	123.601	189.460	147.894	108.690	183.298	20.548	-17.720	59.264
Commuting time (min)	0.562	0.501	0.620	0.483	0.393	0.564	0.096	-0.001	0.191
Rate white	22.374	5.182	39.677	15.356	-4.107	32.631	-1.413	-18.826	16.123
Rate black/African-American	15.544	-6.294	38.285	7.865	-16.566	30.598	-1.699	-23.802	21.181
Rate Asian	44.082	30.186	56.905	40.225	21.799	56.452	6.427	-9.369	22.079
Rate others	117.970	98.319	135.400	103.414	76.883	126.742	-1.395	-24.042	20.813
City population	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table G.11: Regression results for testing H1-2 in 14 areas over 3 months to see whether socio-economic deprivation/inequality indices are correlated to community engagement indicators. DVs are nested within the time.

	Std. Events/Pop			Std. RSVPs/Pop			Std. RSVPs/Event		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H1-3 (HHI): 14 Areas ($N = 280$)									
(Intercept)	-34.375	-55.077	-14.405	-87.514	-104.941	-68.967	-35.627	-48.132	-22.763
Ethnoracial heterogeneity (HHI)	-72.852	-87.851	-58.709	-123.684	-136.641	-109.697	-29.823	-39.233	-20.563
Rate foreign-born	283.853	215.330	354.314	473.081	409.051	532.730	170.038	125.639	212.655
Rate U.S. citizens	5.048	-9.416	19.810	33.526	21.777	44.136	23.548	14.814	32.180
Rate female over 16	-14.432	-30.793	1.843	-32.424	-47.634	-16.446	20.678	9.994	31.525
Median age	0.524	0.398	0.651	0.743	0.652	0.833	0.382	0.307	0.456
Rate adults over 65	-40.704	-51.523	-30.215	-75.854	-83.458	-67.888	-45.817	-52.066	-39.572
Commuting time (min)	0.136	0.101	0.172	0.145	0.120	0.169	0.071	0.051	0.092
Rate white	60.718	42.987	78.974	122.221	100.218	143.351	9.709	-2.987	22.537
Rate black/African-American	80.734	60.741	101.437	154.895	129.609	179.235	14.045	-0.340	28.329
Rate Asian	62.083	43.165	81.899	136.701	112.958	159.367	24.116	10.414	37.855
Rate Hawaiian/Pacific islanders	672.338	554.296	796.873	1210.230	1067.464	1343.655	239.819	154.684	323.592
Rate others	86.960	65.835	108.898	166.101	139.394	191.515	12.962	-2.677	28.448
City population	-0.000	-0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table G.12: Regression results for testing H1-3 to see whether ethnoracial heterogeneity is correlated to community engagement indicators in 14 areas over 20 months. DVs are nested within the time.

	Std. Events/Pop			Std. RSVPs/Pop			Std. RSVPs/Event		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H1-3 (HHI): 28 Areas ($N = 84$)									
(Intercept)	0.933	-22.414	23.525	-1.590	-25.035	21.759	-11.524	-26.399	3.378
Ethnoracial heterogeneity (HHI)	-19.598	-36.821	-2.151	-21.106	-39.526	-2.253	0.675	-10.163	11.227
Rate foreign-born	172.298	117.769	227.856	159.583	102.299	216.470	0.556	-34.404	35.417
Rate U.S. citizens	-0.162	-9.376	8.880	-0.611	-9.761	8.483	2.441	-3.026	7.721
Rate female over 16	23.029	-1.400	46.466	32.930	8.423	56.495	16.451	2.285	30.544
Median age	0.366	0.177	0.549	0.366	0.173	0.563	0.040	-0.071	0.150
Rate adults over 65	-20.477	-33.860	-6.889	-25.345	-38.843	-11.682	-13.239	-21.077	-5.429
Commuting time (min)	0.171	0.106	0.236	0.133	0.063	0.205	0.087	0.048	0.127
Rate white	-17.836	-34.487	-1.230	-17.161	-33.648	-0.059	-1.441	-13.821	10.994
Rate black/African-American	-13.359	-29.990	2.975	-13.271	-29.063	3.566	-3.551	-15.870	8.652
Rate Asian	-20.000	-37.581	-2.332	-16.062	-32.954	1.264	8.563	-4.265	21.498
Rate Hawaiian/Pacific islanders	25.876	-73.498	128.153	10.407	-89.568	112.050	-50.663	-120.416	17.992
Rate others	-9.639	-26.983	7.464	-9.021	-25.985	8.542	-9.268	-22.219	3.669
City population	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000

Table G.13: Regression results for testing H1-3 to see whether ethnoracial heterogeneity is correlated to community engagement indicators in 28 areas over 3 months. DVs are nested within the time.

	Std. Events/Pop			Std. RSVPs/Pop			Std. RSVPs/Event		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H1-3 (HHI): 14 Areas ($N = 42$)									
(Intercept)	1.142	-43.446	47.301	-17.372	-67.187	34.821	-26.450	-57.404	4.747
Ethnoracial heterogeneity (HHI)	-51.685	-84.693	-19.004	-56.510	-91.410	-18.883	-19.089	-41.846	3.775
Rate foreign-born	240.585	79.852	402.012	214.030	42.648	384.241	116.044	5.445	224.800
Rate U.S. citizens	-6.139	-41.297	30.530	0.508	-37.086	40.082	16.130	-8.765	40.184
Rate female over 16	0.948	-33.317	35.872	40.371	2.586	78.197	26.889	3.478	49.603
Median age	0.636	0.305	0.969	0.740	0.363	1.115	0.290	0.064	0.505
Rate adults over 65	-37.825	-65.068	-10.697	-61.574	-92.179	-30.445	-38.205	-56.513	-19.929
Commuting time (min)	0.172	0.080	0.267	0.115	0.009	0.222	0.074	0.013	0.134
Rate white	9.295	-9.913	30.213	6.453	-14.657	26.778	0.686	-16.367	18.242
Rate black/African-American	24.687	2.749	47.545	19.783	-4.092	43.653	1.895	-17.897	21.598
Rate Asian	3.817	-17.509	25.167	8.178	-13.912	30.020	12.637	-6.189	31.292
Rate Hawaiian/Pacific islanders	394.438	219.990	572.369	465.899	263.627	657.984	187.204	48.593	323.671
Rate others	29.726	4.281	54.858	22.695	-3.778	48.895	-2.355	-23.600	18.575
City population	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000	0.000

Table G.14: Regression results for testing H1-3 to see whether ethnoracial heterogeneity is correlated to community engagement indicators in 28 areas over 3 months. DVs are nested within the time.

	Std. Events/Pop			Std. RSVPs/Pop			Std. RSVPs/Event		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H1-4 (HHI*65over): 14 Areas ($N = 280$)									
(Intercept)	-19.996	-50.341	10.011	-36.445	-67.663	-5.355	-40.830	-64.757	-17.118
Ethnoracial heterogeneity (HHI)	40.743	5.893	74.990	51.120	15.079	87.780	28.369	-2.883	60.853
Rate foreign-born	-1.798	-75.658	71.985	-13.033	-90.185	64.603	46.252	-4.841	95.546
Rate U.S. citizens	-16.297	-32.632	0.200	-8.172	-24.718	8.455	16.934	8.277	25.740
Rate female over 16	57.412	37.695	77.039	92.189	71.893	112.439	55.343	41.691	69.427
Median age	0.024	-0.210	0.255	0.077	-0.153	0.308	0.124	-0.077	0.319
Rate adults over 65	146.562	50.734	241.925	173.671	72.058	273.319	74.062	-18.988	168.443
Commuting time (min)	0.251	0.180	0.320	0.285	0.211	0.357	0.141	0.080	0.203
Rate white	-24.690	-33.996	-15.226	-39.726	-48.920	-30.077	-24.276	-29.036	-19.231
Rate black/African-American	-25.406	-34.272	-16.341	-42.140	-51.428	-32.622	-28.354	-33.593	-23.156
Rate Asian	-29.147	-38.264	-19.537	-37.601	-46.989	-28.060	-10.862	-15.968	-5.729
Rate others	-24.386	-34.428	-13.658	-41.142	-52.197	-29.594	-31.930	-37.920	-25.857
City population	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000
HHI:adults over 65	-265.178	-419.251	-110.371	-343.460	-505.332	-179.097	-183.337	-335.185	-34.801

Table G.15: Regression results for testing H1-4 to see whether the effect of ethnoracial heterogeneity on community engagement indicators is moderated by age (percentage of old adults over 65) in 14 areas over 20 months. DVs are nested within the time.

	Std. Events/Pop			Std. RSVPs/Pop			Std. RSVPs/Event		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H1-4 (Gini*65over): 14 Areas ($N = 280$)									
(Intercept)	75.480	56.710	93.942	116.232	104.639	127.694	7.443	-5.695	20.898
Socio-economic inequality (Gini)	14.199	-15.273	43.013	84.184	61.960	106.133	44.898	20.942	69.619
Socio-economic deprivation (depr. index)	-40.087	-45.596	-34.514	-70.638	-74.346	-66.893	-13.295	-17.125	-9.456
Rate foreign-born	-207.726	-257.521	-158.365	-483.224	-515.728	-449.423	-71.974	-106.754	-37.537
Rate female over 16	-208.828	-251.426	-165.143	-408.703	-437.492	-378.979	-68.787	-98.138	-39.180
Median age	-1.982	-2.296	-1.659	-3.395	-3.602	-3.191	-0.271	-0.493	-0.049
Rate adults over 65	169.904	100.067	237.741	394.070	340.328	446.466	95.510	40.135	153.983
Commuting time (min)	0.239	0.204	0.273	0.262	0.243	0.283	0.059	0.034	0.084
Rate white	56.247	42.853	69.924	113.821	104.179	123.307	9.350	-0.140	18.860
Rate black/African-American	69.431	53.793	85.721	139.102	127.849	150.249	12.674	1.559	23.674
Rate Asian	41.269	28.990	54.142	95.786	86.693	104.650	13.234	4.299	22.066
Rate others	92.250	74.872	110.058	167.607	155.252	179.825	7.146	-4.902	19.526
City population	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000
Gini index:adults over 65	37.955	-103.046	179.784	-167.538	-270.624	-64.625	-148.590	-267.292	-32.889

Table G.16: Regression results for testing H1-4 to see whether the effect of socio-economic inequality on community engagement indicators is moderated by age (percentage of old adults over 65) in 14 areas over 20 months. DVs are nested within the time.

	Std. Events/Pop			Std. RSVPs/Pop			Std. RSVPs/Event		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H1-4 (Dprv*65over): 14 Areas ($N = 280$)									
(Intercept)	70.863	57.107	84.184	125.193	117.145	133.094	16.412	7.034	25.761
Socio-economic inequality (Gini)	18.288	7.176	29.875	32.455	25.001	39.857	4.412	-5.020	13.458
Socio-economic deprivation (depr. index)	-51.406	-81.663	-21.317	-123.962	-144.347	-103.526	-42.563	-66.941	-18.971
Rate foreign-born	-228.035	-296.627	-162.185	-555.403	-600.533	-509.375	-104.998	-152.678	-55.568
Rate female over 16	-196.879	-251.193	-143.863	-336.103	-370.542	-301.286	-22.795	-64.357	20.161
Median age	-2.152	-2.734	-1.583	-4.347	-4.738	-3.942	-0.803	-1.238	-0.371
Rate adults over 65	198.345	157.624	240.390	374.493	345.551	402.543	57.955	27.977	87.801
Commuting time (min)	0.284	0.148	0.419	0.516	0.427	0.604	0.205	0.100	0.312
Rate white	55.998	42.999	69.678	109.648	100.241	118.597	5.656	-4.204	15.552
Rate black/African-American	67.542	51.280	83.944	125.860	114.834	136.819	3.531	-8.956	16.329
Rate Asian	41.868	29.546	54.656	95.652	86.807	104.098	11.877	2.746	20.837
Rate others	91.757	74.872	109.506	163.970	151.732	175.566	3.708	-8.883	16.765
City population	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000
Dprv.index:adults over 65	56.900	-98.281	205.704	278.543	178.255	380.641	156.606	36.338	281.888

Table G.17: Regression results for testing H1-4 to see whether the effect of socio-economic deprivation on community engagement indicators is moderated by age (percentage of old adults over 65) in 14 areas over 20 months. DVs are nested within the time.

	Std. Events/Pop			Std. RSVPs/Pop			Std. RSVPs/Event		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H1-4 (HHI*65over): 28 Areas ($N = 84$)									
(Intercept)	3.060	-22.717	29.038	2.925	-23.463	30.401	-12.580	-30.165	5.329
Ethnoracial heterogeneity (HHI)	-17.668	-50.366	14.957	-26.033	-58.650	6.808	-9.143	-31.610	14.121
Rate foreign-born	166.924	112.504	223.868	157.555	99.946	215.415	9.683	-23.855	43.292
Rate U.S. citizens	-0.232	-10.065	9.457	-1.458	-11.742	8.554	1.736	-4.328	7.539
Rate female over 16	23.882	0.494	46.772	33.077	8.678	57.646	15.460	1.176	29.988
Median age	0.359	0.167	0.555	0.372	0.174	0.571	0.072	-0.053	0.195
Rate adults over 65	-13.517	-96.764	70.799	-40.159	-123.365	44.551	-43.780	-103.971	17.259
Commuting time (min)	0.171	0.103	0.240	0.131	0.060	0.203	0.081	0.039	0.124
Rate white	-21.425	-32.345	-10.202	-18.262	-29.571	-6.868	6.094	-1.224	13.334
Rate black/African-American	-16.896	-27.625	-5.517	-14.346	-25.366	-3.144	3.960	-3.335	11.200
Rate Asian	-23.284	-35.731	-10.411	-17.296	-30.433	-4.254	15.553	7.357	23.737
Rate others	-13.185	-25.106	-1.006	-10.334	-22.742	1.819	-1.818	-9.612	5.900
City population	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000
HHI:adults over 65	-11.118	-149.644	123.732	25.144	-113.121	161.170	48.821	-49.527	145.501

Table G.18: Regression results for testing H1-4 to see whether the effect of ethnoracial heterogeneity on community engagement indicators is moderated by age (percentage of old adults over 65) in 28 areas over 3 months. DVs are nested within the time.

	Std. Events/Pop			Std. RSVPs/Pop			Std. RSVPs/Event		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H1-4 (Gini*65over): 28 Areas ($N = 84$)									
(Intercept)	22.438	1.391	42.624	27.313	6.643	48.166	-4.981	-21.638	11.930
Socio-economic inequality (Gini)	-15.469	-49.804	19.404	-25.809	-59.815	9.764	5.216	-23.160	33.247
Socio-economic deprivation (depr. index)	-7.276	-9.596	-4.864	-8.314	-10.686	-5.881	-2.907	-4.566	-1.249
Rate foreign-born	103.901	61.434	146.708	86.816	42.074	131.070	-19.772	-50.704	11.324
Rate female over 16	-15.097	-39.189	9.226	-15.651	-41.425	9.623	-7.450	-25.022	10.265
Median age	-0.099	-0.272	0.081	-0.144	-0.325	0.037	-0.028	-0.153	0.095
Rate adults over 65	2.300	-85.996	90.073	-35.938	-122.849	51.699	-13.946	-86.352	56.223
Commuting time (min)	0.138	0.085	0.191	0.105	0.051	0.157	0.082	0.045	0.119
Rate white	-13.536	-22.308	-4.468	-10.280	-19.218	-0.913	6.101	-0.442	12.639
Rate black/African-American	-10.130	-19.390	-0.895	-7.357	-16.555	2.102	5.068	-1.631	11.747
Rate Asian	-21.029	-30.438	-11.198	-15.579	-25.326	-5.285	13.670	6.444	20.887
Rate others	-5.264	-14.636	4.229	-1.719	-11.049	7.573	-1.023	-7.857	5.832
City population	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000
Gini index:adults over 65	44.378	-142.133	227.159	122.304	-64.051	302.104	18.075	-130.256	169.569

Table G.19: Regression results for testing H1-4 to see whether the effect of socio-economic inequality on community engagement indicators is moderated by age (percentage of old adults over 65) in 28 areas over 3 months. DVs are nested within the time.

	Std. Events/Pop			Std. RSVPs/Pop			Std. RSVPs/Event		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H1-4 (Dprv*65over): 28 Areas ($N = 84$)									
(Intercept)	20.114	7.222	32.462	17.501	4.358	30.097	-6.104	-15.046	2.807
Socio-economic inequality (Gini)	-8.068	-15.695	-0.127	-3.603	-11.678	4.547	8.463	2.932	14.095
Socio-economic deprivation (depr. index)	3.557	-11.426	18.379	-0.912	-16.366	13.974	-0.012	-11.270	11.514
Rate foreign-born	113.854	68.539	157.707	91.763	44.615	137.356	-17.337	-50.189	14.892
Rate female over 16	-20.325	-45.209	5.530	-19.872	-45.401	7.487	-9.425	-27.476	8.963
Median age	-0.003	-0.211	0.212	-0.064	-0.275	0.148	-0.006	-0.160	0.150
Rate adults over 65	19.749	5.650	33.524	18.932	4.833	33.048	-6.027	-16.257	4.274
Commuting time (min)	0.102	0.034	0.171	0.076	0.005	0.147	0.073	0.022	0.123
Rate white	-14.039	-23.098	-4.346	-10.415	-19.696	-1.269	6.226	-0.393	12.856
Rate black/African-American	-10.012	-19.206	-0.335	-7.086	-16.500	2.135	5.381	-1.264	11.935
Rate Asian	-22.575	-32.706	-12.018	-16.544	-26.823	-5.976	13.451	5.949	20.847
Rate others	-5.047	-14.362	4.611	-1.318	-11.030	8.108	-0.645	-7.495	6.065
City population	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000
Dprv.index:adults over 65	-60.699	-141.882	22.625	-42.377	-124.735	41.628	-16.624	-78.990	45.240

Table G.20: Regression results for testing H1-4 to see whether the effect of socio-economic deprivation on community engagement indicators is moderated by age (percentage of old adults over 65) in 28 areas over 3 months. DVs are nested within the time.

	Std. Events/Pop			Std. RSVPs/Pop			Std. RSVPs/Event		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H1-4 (HHI*65over): 14 Areas ($N = 42$)									
(Intercept)	4.932	-53.902	64.001	-16.019	-78.799	45.871	-22.141	-60.980	17.577
Ethnoracial heterogeneity (HHI)	-0.846	-47.451	44.248	2.900	-45.954	50.850	7.861	-29.001	43.594
Rate foreign-born	103.151	-70.212	279.425	62.499	-125.055	257.464	44.112	-64.768	153.346
Rate U.S. citizens	-17.516	-58.930	23.492	-10.794	-54.118	34.774	9.951	-15.063	35.013
Rate female over 16	17.198	-27.435	59.441	56.171	6.928	102.235	40.822	13.830	67.368
Median age	0.217	-0.170	0.619	0.226	-0.201	0.649	0.137	-0.140	0.418
Rate adults over 65	23.704	-79.776	126.330	9.425	-89.913	107.102	-1.476	-97.329	94.652
Commuting time (min)	0.249	0.120	0.373	0.209	0.074	0.347	0.104	0.019	0.191
Rate white	-11.853	-29.657	5.861	-14.888	-33.952	3.738	-17.044	-29.722	-3.580
Rate black/African-American	-8.736	-25.764	8.625	-16.014	-34.628	2.484	-21.619	-34.130	-8.696
Rate Asian	-21.430	-39.205	-3.320	-17.742	-36.829	2.100	-7.150	-20.267	6.175
Rate others	-3.281	-24.519	18.365	-11.467	-34.335	12.202	-26.653	-41.763	-10.783
City population	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000
HHI:adults over 65	-51.287	-215.640	116.085	-55.555	-216.257	106.074	-43.236	-196.236	110.755

Table G.21: Regression results for testing H1-4 to see whether the effect of ethnoracial heterogeneity on community engagement indicators is moderated by age (percentage of old adults over 65) in 14 areas over 3 months. DVs are nested within the time.

	Std. Events/Pop			Std. RSVPs/Pop			Std. RSVPs/Event		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H1-4 (Gini*65over): 14 Areas ($N = 42$)									
(Intercept)	67.359	32.507	99.630	69.562	33.512	103.501	21.486	-6.926	49.289
Socio-economic inequality (Gini)	-9.925	-47.868	28.174	-2.933	-42.385	37.249	10.280	-26.954	46.574
Socio-economic deprivation (depr. index)	-29.577	-38.796	-19.847	-33.939	-42.880	-24.500	-10.999	-18.548	-3.486
Rate foreign-born	-69.050	-157.610	22.472	-166.142	-254.211	-73.237	-65.650	-140.215	10.287
Rate female over 16	-132.635	-204.425	-56.598	-124.030	-195.522	-48.913	-45.076	-103.007	15.036
Median age	-1.311	-1.852	-0.733	-1.390	-1.931	-0.818	-0.315	-0.774	0.128
Rate adults over 65	90.040	2.607	175.478	88.570	0.487	177.207	18.973	-63.365	100.487
Commuting time (min)	0.270	0.189	0.348	0.205	0.117	0.291	0.075	0.011	0.141
Rate white	19.473	-0.237	40.020	15.237	-3.345	34.036	0.767	-15.602	16.964
Rate black/African-American	29.007	5.663	53.357	23.848	1.746	45.957	2.218	-17.045	21.512
Rate Asian	3.660	-14.887	22.675	3.979	-14.368	21.926	5.378	-10.107	20.446
Rate others	50.561	23.863	77.739	42.205	17.185	67.092	-3.405	-24.783	17.581
City population	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000
Gini index:adults over 65	93.879	-88.365	275.482	91.142	-96.901	279.897	11.302	-163.289	190.565

Table G.22: Regression results for testing H1-4 to see whether the effect of socio-economic inequality on community engagement indicators is moderated by age (percentage of old adults over 65) in 14 areas over 3 months. DVs are nested within the time.

	Std. Events/Pop			Std. RSVPs/Pop			Std. RSVPs/Event		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H1-4 (Dprv*65over): 14 Areas ($N = 42$)									
(Intercept)	59.282	31.436	84.082	61.026	31.573	88.645	20.479	-0.984	42.976
Socio-economic inequality (Gini)	11.000	-7.673	29.024	18.981	0.969	36.874	13.064	-2.943	29.249
Socio-economic deprivation (depr. index)	-20.936	-57.634	16.495	-19.706	-55.809	16.960	-9.398	-44.635	26.699
Rate foreign-born	-57.754	-157.187	46.983	-146.271	-252.606	-34.857	-63.926	-152.392	22.202
Rate female over 16	-142.236	-226.602	-53.969	-136.996	-222.764	-50.845	-46.668	-119.371	26.132
Median age	-1.108	-1.863	-0.321	-1.093	-1.849	-0.320	-0.277	-0.963	0.412
Rate adults over 65	120.789	62.548	175.968	111.489	52.498	168.207	21.731	-28.461	70.772
Commuting time (min)	0.224	0.042	0.395	0.137	-0.034	0.310	0.066	-0.102	0.231
Rate white	18.839	-1.984	39.139	14.631	-4.375	33.234	0.537	-15.464	16.539
Rate black/African-American	29.715	4.367	54.704	25.197	1.045	48.648	2.237	-17.791	22.154
Rate Asian	2.479	-17.157	21.241	2.575	-15.818	20.335	5.033	-10.190	19.837
Rate others	49.043	22.490	75.522	40.582	15.392	64.807	-3.748	-24.529	17.189
City population	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000
Dprv.index:adults over 65	-44.144	-231.752	141.284	-71.742	-253.178	109.546	-8.198	-190.856	171.245

Table G.23: Regression results for testing H1-4 to see whether the effect of socio-economic deprivation on community engagement indicators is moderated by age (percentage of old adults over 65) in 14 areas over 3 months. DVs are nested within the time.

	Std. Events/Pop			Std. RSVPs/Pop			Std. RSVPs/Event		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H1-4 (HHI*median_age): 14 Areas ($N = 280$)									
(Intercept)	2.853	-47.042	51.499	-4.433	-56.527	45.565	-24.055	-70.734	22.623
Ethnoracial heterogeneity (HHI)	11.436	-63.222	87.731	8.630	-66.678	89.146	5.433	-70.518	81.065
Rate foreign-born	46.236	-30.061	124.042	51.083	-25.844	129.222	83.248	37.407	130.951
Rate U.S. citizens	-21.045	-37.756	-4.341	-14.171	-31.529	3.421	14.306	5.854	22.910
Rate female over 16	38.415	21.737	55.262	67.530	50.673	84.877	42.554	32.880	52.570
Median age	0.674	-0.344	1.728	0.859	-0.196	1.937	0.551	-0.482	1.567
Rate adults over 65	-16.106	-28.666	-3.625	-37.146	-49.579	-24.407	-39.166	-45.561	-32.941
Commuting time (min)	0.166	0.101	0.233	0.173	0.109	0.242	0.081	0.026	0.133
Rate white	-23.289	-32.817	-14.398	-37.933	-47.617	-28.086	-23.777	-28.612	-18.818
Rate black/African-American	-21.302	-30.205	-12.603	-36.695	-46.038	-27.184	-25.829	-30.655	-21.090
Rate Asian	-31.061	-40.197	-22.022	-40.080	-49.943	-30.248	-12.508	-17.543	-7.411
Rate others	-20.209	-30.457	-9.746	-35.427	-46.559	-24.226	-29.115	-34.761	-23.255
City population	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000
HHI:median age	-0.638	-2.570	1.228	-0.716	-2.726	1.188	-0.384	-2.305	1.530

Table G.24: Regression results for testing H1-4 to see whether the effect of ethnoracial heterogeneity on community engagement indicators is moderated by median age in 14 areas over 20 months. DVs are nested within the time.

	Std. Events/Pop			Std. RSVPs/Pop			Std. RSVPs/Event		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H1-4 (Gini*median_age): 14 Areas ($N = 280$)									
(Intercept)	36.562	1.818	71.431	41.899	13.142	71.343	-9.293	-39.406	21.085
Socio-economic inequality (Gini)	102.482	28.321	173.866	244.916	180.750	307.071	79.652	16.732	143.790
Socio-economic deprivation (depr. index)	-41.050	-46.562	-35.631	-70.093	-73.643	-66.347	-12.546	-16.272	-8.755
Rate foreign-born	-222.432	-272.123	-173.590	-487.179	-519.219	-454.083	-66.100	-100.086	-31.717
Rate female over 16	-217.582	-259.836	-175.007	-409.919	-437.077	-380.393	-65.131	-94.469	-35.461
Median age	-0.936	-1.952	0.063	-0.858	-1.700	-0.027	0.546	-0.341	1.431
Rate adults over 65	185.868	158.176	213.877	308.619	289.762	327.091	25.080	4.578	44.786
Commuting time (min)	0.233	0.201	0.265	0.267	0.249	0.284	0.068	0.046	0.091
Rate white	57.339	43.779	71.203	111.088	101.896	120.042	7.651	-2.002	17.214
Rate black/African-American	70.975	55.381	87.051	136.142	125.303	146.668	10.596	-0.769	21.615
Rate Asian	42.310	29.753	55.025	93.107	84.540	101.443	11.544	2.544	20.541
Rate others	92.516	75.158	110.564	163.088	151.131	174.775	5.261	-7.264	17.761
City population	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000
Gini index:Median age	-2.140	-4.068	-0.208	-5.202	-6.864	-3.496	-1.735	-3.433	-0.050

Table G.25: Regression results for testing H1-4 to see whether the effect of Gini index on community engagement indicators is moderated by median age in 14 areas over 20 months. DVs are nested within the time.

	Std. Events/Pop			Std. RSVPs/Pop			Std. RSVPs/Event		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H1-4 (Dprv*median_age): 14 Areas ($N = 280$)									
(Intercept)	71.256	58.387	83.974	126.310	118.258	134.490	17.376	8.074	26.592
Socio-economic inequality (Gini)	19.136	11.258	27.288	40.928	35.937	45.879	9.526	3.472	15.364
Socio-economic deprivation (depr. index)	-63.244	-93.322	-32.684	-134.692	-156.212	-113.074	-46.606	-69.986	-22.365
Rate foreign-born	-223.715	-270.161	-173.910	-500.278	-533.138	-467.726	-73.998	-108.960	-37.056
Rate female over 16	-203.563	-245.988	-160.439	-373.364	-402.012	-344.829	-45.089	-77.053	-13.262
Median age	-2.061	-2.356	-1.745	-3.655	-3.869	-3.436	-0.410	-0.636	-0.176
Rate adults over 65	192.549	165.107	218.575	328.860	309.721	347.517	32.253	11.723	51.957
Commuting time (min)	0.273	0.213	0.330	0.382	0.343	0.421	0.127	0.081	0.170
Rate white	56.697	43.293	69.637	109.935	100.930	119.052	5.978	-3.962	15.899
Rate black/African-American	68.925	53.241	84.303	131.064	120.424	141.773	6.759	-5.104	18.370
Rate Asian	42.404	30.009	54.401	94.122	85.796	102.649	11.106	1.948	20.321
Rate others	92.435	75.273	109.119	163.282	151.462	175.118	3.529	-9.463	16.280
City population	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000
Dprv.index:median age	0.622	-0.239	1.452	1.793	1.203	2.383	0.951	0.295	1.598

Table G.26: Regression results for testing H1-4 to see whether the effect of Deprivation index on community engagement indicators is moderated by median age in 14 areas over 20 months. DVs are nested within the time.

	Std. Events/Pop			Std. RSVPs/Pop			Std. RSVPs/Event		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H1-4 (HHI*median_age): 28 Areas ($N = 84$)									
(Intercept)	-0.609	-40.851	40.878	2.681	-39.013	43.152	-7.047	-38.824	24.833
Ethnoracial heterogeneity (HHI)	-11.512	-72.010	48.382	-25.884	-85.964	34.014	-18.564	-66.564	29.772
Rate foreign-born	166.113	111.574	221.389	158.732	103.790	214.547	10.153	-23.479	44.040
Rate U.S. citizens	-0.112	-9.475	9.374	-0.942	-10.600	8.808	2.027	-4.117	8.032
Rate female over 16	24.368	0.361	48.180	32.600	8.380	57.559	15.612	1.494	29.982
Median age	0.486	-0.389	1.346	0.285	-0.578	1.149	-0.238	-0.934	0.468
Rate adults over 65	-20.157	-33.405	-6.986	-25.023	-38.163	-11.661	-14.269	-22.294	-6.477
Commuting time (min)	0.172	0.106	0.242	0.131	0.060	0.201	0.081	0.037	0.125
Rate white	-21.763	-32.974	-10.376	-18.108	-29.568	-6.472	6.091	-1.123	13.468
Rate black/African-American	-17.347	-28.576	-5.818	-14.229	-25.508	-3.321	4.022	-3.097	11.287
Rate Asian	-23.670	-36.310	-10.700	-17.046	-29.797	-4.196	15.737	7.441	24.052
Rate others	-13.549	-25.629	-0.815	-10.128	-22.755	2.035	-1.830	-9.668	6.133
City population	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000
HHI:median age	-0.215	-1.712	1.312	0.134	-1.391	1.637	0.503	-0.729	1.719

Table G.27: Regression results for testing H1-4 to see whether the effect of ethnoracial heterogeneity on community engagement indicators is moderated by median age in 28 areas over 3 months. DVs are nested within the time.

	Std. Events/Pop			Std. RSVPs/Pop			Std. RSVPs/Event		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H1-4 (Gini*median_age): 28 Areas ($N = 84$)									
(Intercept)	16.806	-21.606	54.964	20.997	-19.000	59.404	9.308	-25.898	43.618
Socio-economic inequality (Gini)	-3.136	-77.809	74.125	-13.955	-90.235	62.955	-24.530	-94.220	45.692
Socio-economic deprivation (depr. index)	-7.441	-9.709	-5.162	-8.498	-10.919	-6.121	-2.968	-4.561	-1.386
Rate foreign-born	102.314	59.337	143.526	86.510	42.190	131.703	-18.360	-47.289	10.996
Rate female over 16	-15.944	-40.980	8.318	-16.326	-42.703	8.739	-7.879	-25.271	9.222
Median age	-0.044	-1.011	0.985	-0.267	-1.287	0.761	-0.465	-1.397	0.482
Rate adults over 65	23.554	10.377	37.100	21.531	7.729	35.604	-4.875	-14.485	4.585
Commuting time (min)	0.136	0.085	0.187	0.099	0.046	0.153	0.082	0.045	0.119
Rate white	-13.281	-22.205	-4.307	-9.475	-18.885	-0.276	6.255	-0.350	12.915
Rate black/African-American	-9.850	-18.679	-0.737	-6.534	-15.981	2.783	5.313	-1.499	12.099
Rate Asian	-20.785	-30.360	-10.826	-14.830	-25.144	-4.733	13.747	6.532	20.902
Rate others	-5.056	-14.308	4.084	-0.847	-10.232	8.488	-0.662	-7.541	6.247
City population	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000
Gini index:median age	-0.115	-2.239	1.945	0.291	-1.797	2.361	0.909	-1.016	2.830

Table G.28: Regression results for testing H1-4 to see whether the effect of Gini index on community engagement indicators is moderated by median age in 28 areas over 3 months. DVs are nested within the time.

	Std. Events/Pop			Std. RSVPs/Pop			Std. RSVPs/Event		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H1-4 (Dprv*median_age): 28 Areas ($N = 84$)									
(Intercept)	18.800	6.135	31.921	16.104	2.751	29.365	-6.651	-15.382	1.873
Socio-economic inequality (Gini)	-7.397	-15.339	0.588	-3.276	-11.225	4.695	8.481	3.173	13.979
Socio-economic deprivation (depr. index)	-6.462	-30.363	17.084	-7.490	-30.703	16.255	-0.738	-19.959	19.118
Rate foreign-born	103.750	58.261	148.974	85.442	39.779	132.468	-18.681	-50.545	13.764
Rate female over 16	-16.171	-41.913	9.507	-16.409	-43.366	9.744	-7.657	-25.384	10.341
Median age	-0.092	-0.284	0.108	-0.125	-0.316	0.074	-0.019	-0.162	0.119
Rate adults over 65	23.255	8.949	37.260	21.220	7.107	35.349	-5.785	-15.478	3.841
Commuting time (min)	0.134	0.070	0.197	0.099	0.035	0.162	0.078	0.032	0.123
Rate white	-13.270	-22.389	-4.457	-9.666	-19.097	-0.447	6.162	-0.718	13.006
Rate black/African-American	-9.810	-18.893	-0.863	-6.761	-16.162	2.656	5.158	-1.653	12.165
Rate Asian	-20.843	-30.776	-10.929	-15.024	-25.471	-5.028	13.659	6.000	21.204
Rate others	-5.011	-14.459	4.250	-1.196	-11.103	8.763	-1.018	-8.047	6.101
City population	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000
Dprv.index:median age	-0.027	-0.671	0.640	-0.028	-0.693	0.627	-0.062	-0.615	0.475

Table G.29: Regression results for testing H1-4 to see whether the effect of Deprivation index on community engagement indicators is moderated by median age in 28 areas over 3 months. DVs are nested within the time.

	Std. Events/Pop			Std. RSVPs/Pop			Std. RSVPs/Event		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H1-4 (HHI*median_age): 14 Areas ($N = 42$)									
(Intercept)	10.820	-52.066	73.545	-7.992	-79.323	63.561	-15.492	-67.243	37.432
Ethnoracial heterogeneity (HHI)	-6.860	-83.933	67.986	-7.843	-87.033	69.667	-2.390	-76.624	70.115
Rate foreign-born	109.344	-73.745	285.681	78.389	-104.107	254.526	56.269	-59.035	165.721
Rate U.S. citizens	-19.107	-60.160	20.977	-11.528	-53.892	32.228	9.686	-16.144	34.999
Rate female over 16	13.179	-29.153	53.941	49.530	2.784	95.011	37.269	11.863	61.906
Median age	0.316	-0.706	1.335	0.274	-0.811	1.342	0.170	-0.818	1.158
Rate adults over 65	-6.724	-33.549	21.103	-23.891	-54.467	7.349	-28.450	-45.946	-9.888
Commuting time (min)	0.231	0.106	0.355	0.187	0.050	0.324	0.086	0.007	0.169
Rate white	-11.751	-28.691	6.197	-13.976	-33.150	5.378	-16.995	-29.905	-3.614
Rate black/African-American	-8.216	-24.843	8.275	-14.527	-32.792	3.257	-20.987	-33.249	-8.277
Rate Asian	-22.222	-40.337	-3.692	-17.814	-37.710	1.522	-7.611	-20.876	5.517
Rate others	-2.961	-24.135	19.128	-9.943	-32.521	13.591	-25.799	-40.621	-10.860
City population	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000
HHI:median age (HHI)	-0.095	-1.950	1.790	0.013	-1.912	1.944	0.035	-1.769	1.876

Table G.30: Regression results for testing H1-4 to see whether the effect of ethnoracial heterogeneity on community engagement indicators is moderated by median age in 14 areas over 3 months. DVs are nested within the time.

	Std. Events/Pop			Std. RSVPs/Pop			Std. RSVPs/Event		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H1-4 (Gini*median_age): 14 Areas ($N = 42$)									
(Intercept)	42.007	-0.023	84.070	42.407	-1.158	84.505	17.156	-21.701	55.741
Socio-economic inequality (Gini)	47.038	-33.119	124.997	52.059	-25.444	130.403	20.700	-52.885	96.336
Socio-economic deprivation (depr. index)	-30.246	-39.580	-20.194	-33.789	-43.136	-23.925	-11.258	-18.932	-3.775
Rate foreign-born	-79.086	-168.090	17.181	-171.362	-266.419	-68.864	-69.007	-143.278	5.173
Rate female over 16	-138.364	-209.138	-62.649	-122.665	-194.479	-48.811	-46.698	-105.473	12.080
Median age	-0.776	-1.895	0.299	-0.830	-1.881	0.217	-0.209	-1.236	0.825
Rate adults over 65	130.768	82.308	176.958	124.431	76.810	169.418	24.178	-13.895	62.085
Commuting time (min)	0.261	0.185	0.332	0.197	0.110	0.281	0.073	0.010	0.139
Rate white	19.899	-0.450	40.581	14.256	-4.616	32.938	0.870	-15.643	17.560
Rate black/African-American	29.791	6.073	54.028	22.720	0.739	44.928	2.456	-16.773	21.743
Rate Asian	4.072	-15.374	23.756	3.294	-15.223	21.448	5.426	-10.174	21.253
Rate others	50.078	23.573	76.862	39.644	15.387	64.092	-3.328	-24.867	18.389
City population	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000
Gini index:median age	-1.017	-3.053	1.068	-1.003	-3.031	1.042	-0.213	-2.244	1.729

Table G.31: Regression results for testing H1-4 to see whether the effect of Gini index on community engagement indicators is moderated by median age in 14 areas over 3 months. DVs are nested within the time.

	Std. Events/Pop			Std. RSVPs/Pop			Std. RSVPs/Event		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H1-4 (Dprv*median_age): 14 Areas ($N = 42$)									
(Intercept)	57.865	31.384	82.295	58.824	29.553	87.046	19.874	-1.369	41.240
Socio-economic inequality (Gini)	8.582	-5.820	22.398	15.015	-0.160	29.626	12.347	0.254	24.444
Socio-economic deprivation (depr. index)	-30.071	-69.509	12.710	-28.613	-69.359	11.820	-10.418	-47.534	27.364
Rate foreign-born	-70.052	-153.509	21.342	-160.195	-255.334	-59.805	-65.543	-138.685	9.970
Rate female over 16	-132.283	-200.826	-58.826	-120.179	-194.243	-45.975	-43.419	-103.962	16.778
Median age	-1.243	-1.766	-0.701	-1.271	-1.807	-0.679	-0.296	-0.731	0.139
Rate adults over 65	129.088	82.298	173.158	121.971	74.010	167.883	22.585	-14.939	59.320
Commuting time (min)	0.262	0.158	0.359	0.189	0.081	0.295	0.072	-0.017	0.157
Rate white	18.801	-0.714	38.018	13.411	-4.821	32.033	0.219	-15.773	16.747
Rate black/African-American	28.366	5.196	51.101	21.949	0.203	44.120	1.602	-17.499	21.646
Rate Asian	3.084	-15.338	21.333	2.311	-15.232	19.951	4.946	-10.075	20.206
Rate others	49.017	23.679	74.525	38.990	14.983	63.785	-4.244	-25.120	17.381
City population	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000
Dprv.index:Median age	0.017	-1.109	1.065	-0.124	-1.197	0.997	-0.011	-1.029	0.988

Table G.32: Regression results for testing H1-4 to see whether the effect of Deprivation index on community engagement indicators is moderated by median age in 14 areas over 3 months. DVs are nested within the time.

	Std. Events/Pop			Std. RSVPs/Pop			Std. RSVPs/Event		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H1-4 (HHI*gender): 14 Areas ($N = 280$)									
(Intercept)	53.867	-0.207	105.309	58.667	3.151	112.683	10.951	-37.800	61.422
Ethnoracial heterogeneity (HHI)	-74.844	-156.065	8.902	-98.269	-183.094	-12.235	-53.611	-136.902	27.224
Rate foreign-born	60.367	-10.212	129.712	69.541	-2.367	137.959	92.106	56.002	128.051
Rate U.S. citizens	-20.872	-37.402	-3.871	-13.701	-30.455	2.723	14.382	5.659	22.974
Rate female over 16	-40.356	-144.571	64.353	-34.170	-139.876	73.884	-13.202	-118.393	87.606
Median age	0.351	0.197	0.502	0.506	0.351	0.660	0.362	0.279	0.441
Rate adults over 65	-16.601	-28.988	-4.151	-38.029	-50.364	-25.717	-39.723	-46.019	-33.315
Commuting time (min)	0.148	0.105	0.191	0.150	0.108	0.193	0.068	0.047	0.090
Rate white	-21.871	-31.272	-12.413	-36.224	-45.907	-26.319	-22.951	-28.478	-17.127
Rate black/African-American	-19.689	-28.696	-10.906	-34.873	-44.001	-25.850	-24.841	-29.923	-19.523
Rate Asian	-30.110	-39.710	-20.887	-38.984	-48.502	-29.432	-11.921	-17.353	-6.281
Rate others	-17.774	-28.508	-7.562	-32.579	-43.499	-21.897	-27.630	-33.719	-21.476
City population	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000
HHI:female over 16	123.078	-42.832	288.113	158.993	-12.002	325.806	87.583	-75.426	255.240

Table G.33: Regression results for testing H1-4 to see whether the effect of ethnoracial heterogeneity on community engagement indicators is moderated by gender balance (female %) in 14 areas over 20 months. DVs are nested within the time.

	Std. Events/Pop			Std. RSVPs/Pop			Std. RSVPs/Event		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H1-4 (Gini*gender): 14 Areas ($N = 280$)									
(Intercept)	127.329	83.304	170.808	258.610	218.905	298.915	53.745	12.050	95.651
Socio-economic inequality (Gini)	-100.978	-196.686	-8.024	-245.825	-336.712	-154.163	-63.081	-158.559	29.510
Socio-economic deprivation (depr. index)	-39.365	-44.545	-33.854	-65.085	-68.823	-61.365	-11.106	-15.139	-6.908
Rate foreign-born	-207.986	-254.291	-159.775	-444.955	-476.621	-413.035	-53.146	-88.577	-17.491
Rate female over 16	-312.362	-399.405	-227.641	-628.248	-703.268	-554.566	-121.256	-200.902	-41.376
Median age	-1.920	-2.221	-1.609	-3.203	-3.418	-2.991	-0.248	-0.478	-0.008
Rate adults over 65	183.037	155.170	209.802	297.395	278.375	316.558	22.302	0.836	42.545
Commuting time (min)	0.246	0.213	0.279	0.297	0.278	0.317	0.076	0.052	0.099
Rate white	53.511	39.501	66.799	99.148	89.850	109.134	4.387	-6.345	14.727
Rate black/African-American	66.327	49.838	81.802	121.746	110.641	133.266	6.625	-5.895	18.621
Rate Asian	38.552	25.513	50.752	81.556	72.649	90.903	8.399	-1.631	18.016
Rate others	89.573	71.715	106.688	152.551	140.371	164.944	2.462	-11.235	15.508
City population	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000
Rate female over 16	234.163	56.899	413.996	559.639	389.069	730.661	147.011	-27.956	327.020

Table G.34: Regression results for testing H1-4 to see whether the effect of Gini index on community engagement indicators is moderated by gender balance (female %) in 14 areas over 20 months. DVs are nested within the time.

	Std. Events/Pop			Std. RSVPs/Pop			Std. RSVPs/Event		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H1-4 (Dprv*gender): 14 Areas ($N = 280$)									
(Intercept)	71.874	57.938	85.036	129.702	120.529	138.834	21.272	11.700	30.658
Socio-economic inequality (Gini)	22.425	15.634	29.259	49.281	44.816	53.710	14.507	9.744	19.150
Socio-economic deprivation (depr. index)	-23.261	-66.810	19.574	-77.791	-116.984	-39.225	-30.699	-71.931	8.592
Rate foreign-born	-210.758	-258.557	-160.462	-465.570	-500.699	-431.050	-63.164	-97.700	-27.494
Rate female over 16	-213.600	-255.398	-169.707	-394.818	-424.571	-365.931	-61.320	-91.300	-30.410
Median age	-1.942	-2.254	-1.604	-3.430	-3.690	-3.176	-0.366	-0.609	-0.110
Rate adults over 65	184.804	154.933	213.078	316.339	293.938	339.629	32.275	10.107	53.664
Commuting time (min)	0.237	0.203	0.270	0.275	0.255	0.294	0.070	0.048	0.092
Rate white	57.631	43.736	70.362	110.173	100.560	120.168	7.735	-2.018	17.446
Rate black/African-American	71.031	54.619	86.037	134.559	123.366	146.406	10.614	-0.779	21.825
Rate Asian	42.975	29.902	55.261	92.097	83.127	101.326	11.200	1.838	20.382
Rate others	93.078	74.888	109.345	164.423	151.809	177.667	6.844	-5.870	19.190
City population	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000
Dprv.index:female over 16	-32.473	-111.245	47.763	16.734	-54.824	89.324	34.244	-38.338	110.279

Table G.35: Regression results for testing H1-4 to see whether the effect of Deprivation index on community engagement indicators is moderated by gender balance (female %) in 14 areas over 20 months. DVs are nested within the time.

	Std. Events/Pop			Std. RSVPs/Pop			Std. RSVPs/Event		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H1-4 (HHI*gender): 28 Areas ($N = 84$)									
(Intercept)	15.367	-45.590	76.180	20.870	-39.688	80.071	-5.952	-62.213	50.375
Ethnoracial heterogeneity (HHI)	-37.614	-136.404	59.267	-55.906	-150.709	40.694	-21.116	-109.914	72.035
Rate foreign-born	166.732	110.634	223.067	154.885	99.956	210.241	6.693	-26.429	39.159
Rate U.S. citizens	-0.636	-9.667	8.350	-0.900	-10.105	8.109	3.056	-2.323	8.524
Rate female over 16	2.627	-116.702	120.913	-9.790	-125.083	107.366	-10.919	-121.854	102.251
Median age	0.366	0.183	0.552	0.364	0.176	0.555	0.046	-0.064	0.155
Rate adults over 65	-20.259	-33.110	-6.901	-24.744	-38.637	-11.231	-13.902	-21.918	-6.032
Commuting time (min)	0.168	0.101	0.235	0.133	0.067	0.201	0.087	0.047	0.127
Rate white	-21.282	-33.176	-10.052	-17.976	-29.036	-6.436	6.332	-1.143	13.669
Rate black/African-American	-16.777	-27.955	-5.684	-14.346	-25.435	-3.185	3.938	-3.314	11.086
Rate Asian	-23.182	-36.276	-10.235	-16.868	-29.487	-3.814	16.248	7.511	24.829
Rate others	-13.149	-25.620	-0.947	-10.132	-22.059	2.379	-1.536	-9.645	6.462
City population	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000
HHI:female over 16	34.672	-153.201	225.551	69.603	-118.572	254.621	43.242	-139.212	220.652

Table G.36: Regression results for testing H1-4 to see whether the effect of ethnoracial heterogeneity on community engagement indicators is moderated by gender balance (female %) in 28 areas over 3 months. DVs are nested within the time.

	Std. Events/Pop			Std. RSVPs/Pop			Std. RSVPs/Event		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H1-4 (Gini*gender): 28 Areas ($N = 84$)									
(Intercept)	20.047	-36.498	75.203	28.331	-28.406	85.533	-15.495	-68.570	37.792
Socio-economic inequality (Gini)	-10.703	-127.197	107.962	-30.571	-149.492	88.308	27.536	-86.304	141.179
Socio-economic deprivation (depr. index)	-7.398	-9.692	-5.047	-8.417	-10.757	-5.962	-3.015	-4.670	-1.334
Rate foreign-born	102.718	59.727	144.562	84.083	38.669	129.683	-19.455	-50.045	10.811
Rate female over 16	-18.411	-125.723	89.359	-39.985	-151.861	68.439	9.415	-94.128	113.274
Median age	-0.095	-0.265	0.074	-0.128	-0.301	0.051	-0.030	-0.158	0.096
Rate adults over 65	23.226	9.898	36.698	21.200	7.232	35.012	-5.039	-14.725	4.566
Commuting time (min)	0.135	0.085	0.188	0.102	0.048	0.157	0.080	0.042	0.116
Rate white	-13.281	-22.211	-4.187	-9.519	-18.636	-0.588	6.384	-0.198	13.059
Rate black/African-American	-9.857	-18.820	-0.573	-6.684	-15.973	2.375	5.362	-1.352	12.106
Rate Asian	-20.787	-30.424	-11.025	-14.755	-24.933	-5.002	13.904	6.589	21.080
Rate others	-5.003	-14.277	4.437	-0.995	-10.471	8.175	-0.725	-7.504	6.308
City population	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000
Rate female over 16	6.334	-222.315	230.817	52.158	-177.963	280.681	-36.657	-256.062	183.715

Table G.37: Regression results for testing H1-4 to see whether the effect of Gini index on community engagement indicators is moderated by gender balance (female %) in 28 areas over 3 months. DVs are nested within the time.

	Std. Events/Pop			Std. RSVPs/Pop			Std. RSVPs/Event		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H1-4 (Dprv*gender): 28 Areas ($N = 84$)									
(Intercept)	18.995	6.116	31.372	16.819	3.676	29.603	-6.489	-15.024	2.305
Socio-economic inequality (Gini)	-7.370	-14.889	-0.059	-3.014	-10.967	5.041	8.505	2.862	13.966
Socio-economic deprivation (depr. index)	0.777	-30.893	32.004	3.376	-27.380	34.345	0.603	-29.788	30.350
Rate foreign-born	100.655	58.571	143.477	80.745	35.786	125.925	-21.688	-53.351	9.753
Rate female over 16	-16.795	-42.462	8.589	-18.802	-45.458	7.585	-8.044	-26.401	10.239
Median age	-0.093	-0.268	0.083	-0.128	-0.312	0.058	-0.028	-0.153	0.100
Rate adults over 65	23.239	9.578	36.500	21.487	7.003	35.658	-5.207	-15.099	4.667
Commuting time (min)	0.136	0.084	0.187	0.101	0.047	0.153	0.082	0.046	0.118
Rate white	-13.185	-22.180	-4.058	-9.239	-18.443	-0.021	6.369	-0.121	12.843
Rate black/African-American	-9.674	-18.746	-0.287	-6.229	-15.650	3.143	5.362	-1.399	11.833
Rate Asian	-20.555	-30.563	-10.582	-14.406	-24.525	-4.142	13.992	6.907	20.992
Rate others	-4.895	-14.141	4.690	-0.672	-10.200	8.643	-0.735	-7.366	5.966
City population	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000
Rate female over 16	-15.798	-75.657	44.744	-23.115	-82.546	35.978	-6.907	-64.386	51.577

Table G.38: Regression results for testing H1-4 to see whether the effect of Deprivation index on community engagement indicators is moderated by gender balance (female %) in 28 areas over 3 months. DVs are nested within the time.

	Std. Events/Pop			Std. RSVPs/Pop			Std. RSVPs/Event		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H1-4 (HHI*gender): 14 Areas ($N = 42$)									
Ethnoracial heterogeneity (HHI)	-20.660	-106.025	66.789	-20.086	-105.582	66.163	-16.463	-97.222	65.528
Rate foreign-born	117.398	-52.524	291.415	73.140	-111.466	258.051	54.124	-52.450	157.231
Rate U.S. citizens	-18.303	-57.846	23.103	-12.471	-56.239	31.605	9.092	-15.780	33.467
Rate female over 16	-1.055	-110.675	110.948	33.912	-78.335	146.372	18.626	-83.100	124.669
Median age	0.287	-0.086	0.655	0.279	-0.136	0.672	0.196	-0.026	0.415
Rate adults over 65	-7.734	-36.287	20.704	-23.222	-53.794	8.640	-28.366	-45.997	-9.856
Commuting time (min)	0.225	0.114	0.334	0.190	0.066	0.317	0.086	0.025	0.147
Rate white	-11.998	-29.477	5.696	-13.347	-32.238	6.308	-16.922	-29.550	-3.407
Rate black/African-American	-8.510	-26.245	7.786	-14.235	-32.528	3.927	-20.966	-32.828	-8.033
Rate Asian	-22.689	-41.177	-4.936	-17.435	-37.197	2.237	-7.639	-19.907	5.689
Rate others	-2.912	-24.429	18.367	-9.684	-33.358	13.538	-25.913	-40.226	-10.660
City population	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000
HHI:female over 16	21.403	-147.523	187.260	26.584	-139.253	193.496	30.396	-135.365	191.015

Table G.39: Regression results for testing H1-4 to see whether the effect of ethnoracial heterogeneity on community engagement indicators is moderated by gender balance (female %) in 14 areas over 3 months. DVs are nested within the time.

	Std. Events/Pop			Std. RSVPs/Pop			Std. RSVPs/Event		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H1-4 (Gini*gender): 14 Areas ($N = 42$)									
(Intercept)	80.589	24.328	134.344	78.069	21.418	133.590	25.606	-23.493	75.346
Socio-economic inequality (Gini)	-38.041	-136.551	61.748	-25.261	-125.839	80.053	1.100	-102.338	101.601
Socio-economic deprivation (depr. index)	-29.743	-38.408	-20.268	-33.586	-43.038	-23.830	-11.009	-18.440	-3.574
Rate foreign-born	-73.971	-159.274	17.011	-167.855	-262.471	-68.990	-67.868	-141.447	4.149
Rate female over 16	-176.733	-293.032	-55.106	-157.495	-278.255	-35.128	-54.853	-157.656	46.942
Median age	-1.250	-1.758	-0.707	-1.311	-1.848	-0.763	-0.309	-0.737	0.119
Rate adults over 65	130.285	83.434	174.026	125.269	78.475	171.889	23.869	-13.450	60.748
Commuting time (min)	0.265	0.190	0.338	0.201	0.113	0.288	0.076	0.011	0.139
Rate white	18.955	-1.079	38.176	14.215	-4.388	33.898	0.570	-15.810	16.555
Rate black/African-American	28.680	5.655	50.765	22.669	0.582	46.048	2.002	-17.016	21.079
Rate Asian	3.076	-15.745	20.991	3.110	-14.976	21.745	5.203	-10.423	20.853
Rate others	49.740	24.234	75.180	40.371	15.366	66.558	-3.582	-24.397	17.424
City population	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000
Rate female over 16	90.404	-104.310	279.497	77.133	-124.811	272.522	21.789	-167.901	215.237

Table G.40: Regression results for testing H1-4 to see whether the effect of Gini index on community engagement indicators is moderated by gender balance (female %) in 14 areas over 3 months. DVs are nested within the time.

	Std. Events/Pop			Std. RSVPs/Pop			Std. RSVPs/Event		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H1-4 (Dprv*gender): 14 Areas ($N = 42$)									
(Intercept)	57.471	30.368	83.521	59.478	31.425	85.855	20.273	-1.393	41.318
Socio-economic inequality (Gini)	8.787	-5.106	21.911	15.337	0.892	28.788	12.691	2.170	23.501
Socio-economic deprivation (depr. index)	-9.888	-53.608	31.856	-13.749	-54.935	27.052	-7.065	-49.066	35.476
Rate foreign-born	-69.220	-156.231	24.499	-166.081	-258.045	-69.604	-67.553	-141.587	8.750
Rate female over 16	-133.744	-208.481	-59.077	-124.952	-196.230	-49.064	-45.534	-102.991	14.010
Median age	-1.198	-1.740	-0.647	-1.273	-1.805	-0.724	-0.303	-0.739	0.141
Rate adults over 65	125.465	78.129	172.982	122.314	73.617	168.932	23.247	-15.160	60.867
Commuting time (min)	0.261	0.183	0.338	0.196	0.109	0.279	0.075	0.010	0.141
Rate white	18.922	-0.502	40.090	14.963	-4.154	34.158	0.874	-15.456	17.190
Rate black/African-American	28.480	5.308	53.405	23.587	0.771	46.261	2.340	-16.888	21.481
Rate Asian	3.611	-15.085	23.684	4.239	-14.161	22.888	5.605	-10.062	21.004
Rate others	48.489	22.919	76.137	40.421	15.019	65.338	-3.499	-24.951	17.736
City population	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000
Dprv.index:female over 16	-37.016	-113.645	43.866	-37.872	-114.734	38.683	-7.565	-84.979	70.254

Table G.41: Regression results for testing H1-4 to see whether the effect of Deprivation index on community engagement indicators is moderated by gender balance (female %) in 14 areas over 3 months. DVs are nested within the time.

	Std. Events/Pop			Std. RSVPs/Pop			Std. RSVPs/Event		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H1-5 (HHI) ctl. SES: 14 Areas ($N = 280$)									
(Intercept)	82.840	68.125	97.640	145.570	136.752	153.923	27.842	17.835	37.692
Ethnoracial heterogeneity (HHI)	-13.335	-22.783	-3.804	-19.391	-24.304	-14.505	-11.181	-17.258	-5.103
Rate foreign-born	-219.258	-267.586	-171.990	-479.548	-508.719	-448.469	-61.249	-94.923	-27.000
Socio-economic inequality (Gini)	23.229	16.599	30.072	51.743	47.628	55.668	15.098	10.236	19.741
Socio-economic deprivation (depr. index)	-41.159	-46.561	-35.960	-70.393	-73.720	-66.924	-12.292	-16.136	-8.467
Rate female over 16	-213.854	-254.817	-173.704	-403.300	-429.436	-375.304	-59.410	-89.399	-28.083
Median age	-1.991	-2.303	-1.687	-3.453	-3.646	-3.254	-0.290	-0.512	-0.067
Rate adults over 65	187.870	161.702	215.266	317.165	299.269	334.425	24.852	5.146	45.088
Commuting time (min)	0.235	0.201	0.267	0.274	0.256	0.291	0.069	0.047	0.090
Rate white	54.788	41.916	68.122	108.654	99.618	117.219	4.562	-5.518	14.384
Rate black/African-American	71.555	56.569	87.126	138.269	127.664	148.047	9.974	-1.589	21.520
Rate Asian	42.207	30.453	54.462	94.179	85.735	102.116	10.609	1.386	19.681
Rate others	94.312	77.664	111.642	168.279	156.671	179.262	5.513	-7.224	18.315
City population	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000

Table G.42: Regression results for testing H1-5 to see whether the socio-economic deprivation/inequality mediates the effect of ethnoraical heterogeneity on community engagement indicators in 14 areas over 20 months. DVs are nested within the time.

	Std. Events/Pop			Std. RSVPs/Pop			Std. RSVPs/Event		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H1-5 (HHI) ctl. SES: 28 Areas ($N = 84$)									
(Intercept)	20.743	7.614	33.700	17.778	3.972	31.756	-8.690	-17.628	0.547
Ethnoracial heterogeneity (HHI)	-8.156	-23.625	6.530	-6.604	-21.881	8.972	6.854	-4.124	17.775
Rate foreign-born	111.282	66.931	154.724	91.450	45.654	138.964	-26.164	-56.867	5.499
Socio-economic inequality (Gini)	-8.198	-16.000	-0.604	-3.969	-12.181	4.270	9.109	3.706	14.627
Socio-economic deprivation (depr. index)	-6.966	-9.379	-4.601	-8.162	-10.615	-5.662	-3.287	-5.028	-1.545
Rate female over 16	-9.951	-37.116	17.145	-11.714	-39.527	16.549	-11.942	-30.374	7.527
Median age	-0.057	-0.244	0.125	-0.100	-0.294	0.091	-0.057	-0.193	0.079
Rate adults over 65	19.144	4.250	34.846	18.073	1.990	34.080	-1.980	-13.048	8.811
Commuting time (min)	0.140	0.086	0.194	0.104	0.051	0.156	0.077	0.042	0.115
Rate white	-14.301	-23.406	-4.976	-10.321	-19.357	-1.215	7.127	0.346	13.746
Rate black/African-American	-9.578	-18.850	-0.266	-6.387	-15.194	2.772	4.970	-1.914	11.578
Rate Asian	-20.148	-30.007	-9.953	-14.280	-23.879	-4.252	13.259	5.850	20.190
Rate others	-4.644	-13.912	4.939	-0.731	-9.967	8.544	-1.159	-8.043	5.589
City population	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000

Table G.43: Regression results for testing H1-5 to see whether the socio-economic deprivation/inequality mediates the effect of ethnoraical heterogeneity on community engagement indicators in 28 areas over 3 months. DVs are nested within the time.

	Std. Events/Pop			Std. RSVPs/Pop			Std. RSVPs/Event		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H1-5 (HHI) ctl. SES: 14 Areas ($N = 42$)									
(Intercept)	72.750	40.430	102.789	75.971	41.468	108.288	23.290	-1.772	48.276
Ethnoracial heterogeneity (HHI)	-18.137	-38.920	3.135	-19.901	-44.354	4.823	-4.560	-23.447	14.329
Rate foreign-born	-83.545	-174.935	15.536	-179.064	-272.237	-77.415	-69.080	-140.478	3.100
Socio-economic inequality (Gini)	10.520	-3.741	23.342	16.882	2.487	30.705	12.741	2.160	23.450
Socio-economic deprivation (depr. index)	-30.815	-40.348	-20.721	-34.896	-44.522	-25.002	-11.134	-18.293	-3.864
Rate female over 16	-138.155	-211.579	-60.558	-127.536	-200.530	-53.264	-44.340	-101.184	13.162
Median age	-1.283	-1.818	-0.732	-1.345	-1.881	-0.750	-0.310	-0.731	0.106
Rate adults over 65	131.266	83.111	178.378	126.597	75.443	172.953	23.320	-13.227	59.927
Commuting time (min)	0.261	0.184	0.336	0.195	0.109	0.279	0.075	0.010	0.138
Rate white	16.854	-4.339	37.709	11.514	-8.142	31.894	-0.144	-16.081	15.875
Rate black/African-American	31.227	6.641	55.754	25.256	2.472	49.862	2.388	-15.803	20.998
Rate Asian	4.387	-15.222	23.988	3.997	-14.444	23.051	5.376	-9.600	20.539
Rate others	52.298	24.851	79.518	43.008	16.800	70.372	-3.229	-23.277	17.662
City population	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000

Table G.44: Regression results for testing H1-5 to see whether the socio-economic deprivation/inequality mediates the effect of ethnoraical heterogeneity on community engagement indicators in 14 areas over 3 months. DVs are nested within the time.

	Model w/o Ethnic			Model w/ Ethnic		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H2 (Dprv-Frag):14 Areas ($N = 280$)						
(Intercept)	3.008	-2.415	8.279	-22.155	-35.483	-9.102
Socio-economic inequality (Gini)	-2.183	-6.114	1.858	-10.865	-17.862	-3.979
Socio-economic deprivation (depr. index)	-4.796	-6.429	-3.140	-11.811	-17.540	-6.062
Rate U.S. citizens	1.733	-0.704	4.121	43.685	25.419	61.701
Rate female over 16	-4.150	-11.528	2.962	13.073	-22.468	47.908
Median age	-0.054	-0.180	0.075	-0.656	-1.032	-0.275
Rate adults over 65	12.829	2.788	22.743	40.416	13.834	67.604
Commuting time (min)	-0.080	-0.106	-0.054	-0.019	-0.074	0.034
City population	0.000	0.000	0.000	0.000	0.000	0.000
Rate white				-4.657	-20.788	11.635
Rate black/African-American				-7.272	-24.679	10.471
Rate Asian				7.553	-8.038	22.789
Rate Hawaiian/Pacific islanders				-97.979	-189.913	-2.325
Rate others				6.589	-12.228	25.623

Table G.45: Regression results for testing H2 to see whether socio-economic inequality/deprivation indices are correlated to the fragmentation of local information in 14 areas over 20 months. DVs are nested within the time.

	Model w/o Ethnic			Model w/ Ethnic		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H2 (Dprv-Frag):28 Areas ($N = 84$)						
(Intercept)	24.714	14.816	34.351	4.955	-19.716	30.972
Socio-economic inequality (Gini)	-21.426	-33.062	-9.973	-27.922	-40.470	-15.514
Socio-economic deprivation (depr. index)	1.860	-0.676	4.341	5.971	2.659	9.324
Rate U.S. citizens	-9.707	-16.082	-3.288	-4.504	-14.888	5.595
Rate female over 16	4.676	-16.297	25.644	57.667	19.918	95.672
Median age	-0.310	-0.550	-0.077	-0.262	-0.519	-0.019
Rate adults over 65	38.095	20.868	55.469	22.453	3.174	42.366
Commuting time (min)	-0.146	-0.227	-0.068	-0.129	-0.212	-0.048
City population	-0.000	-0.000	0.000	0.000	-0.000	0.000
Rate white				-8.678	-26.623	9.846
Rate black/African-American				-12.145	-30.107	6.144
Rate Asian				-1.627	-20.159	17.534
Rate Hawaiian/Pacific islanders				-29.299	-146.953	91.169
Rate others				-13.530	-32.317	5.644

Table G.46: Regression results for testing H2 sto see whether socio-economic inequality/deprivation indices are correlated to the fragmentation of local information in 28 areas over 3 months. DVs are nested within the time.

	Model w/o Ethnic			Model w/ Ethnic		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H2 (Dprv-Frag):14 Areas ($N = 42$)						
(Intercept)	13.540	-9.374	36.017	-21.615	-68.791	27.263
Socio-economic inequality (Gini)	-11.463	-28.534	5.938	-20.644	-43.840	3.328
Socio-economic deprivation (depr. index)	-8.043	-14.758	-1.206	-14.006	-28.986	0.756
Rate U.S. citizens	-5.730	-15.983	4.677	42.048	-5.592	90.202
Rate female over 16	-0.653	-31.709	30.106	24.657	-75.451	123.286
Median age	-0.135	-0.649	0.392	-0.675	-1.604	0.280
Rate adults over 65	28.919	-12.153	68.567	46.123	-20.834	112.572
Commuting time (min)	-0.132	-0.241	-0.024	-0.090	-0.268	0.086
City population	0.000	-0.000	0.000	0.000	0.000	0.000
Rate white				-4.007	-24.014	16.434
Rate black/African-American				-6.556	-29.561	16.089
Rate Asian				12.040	-9.208	33.190
Rate Hawaiian/Pacific islanders				-83.870	-326.929	157.950
Rate others				5.118	-23.143	32.746

Table G.47: Regression results for testing H2 sto see whether socio-economic inequality/deprivation indices are correlated to the fragmentation of local information in 14 areas over 3 months. DVs are nested within the time.

	Std. Events/Pop			Std. RSVPs/Pop			Std. RSVPs/Event		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H3 (1) (Frag-Engage): 14 Areas ($N = 280$)									
(Intercept)	-29.093	-52.017	-6.932	-53.453	-75.596	-31.003	-28.561	-41.548	-15.627
The fragmentation of info (β_{CO})	0.327	-3.246	3.830	3.377	1.292	5.506	1.050	-0.214	2.290
Rate foreign-born	162.553	88.257	237.179	209.704	134.008	285.881	109.564	68.231	150.710
Rate U.S. citizens	-8.591	-25.294	7.950	3.203	-13.630	20.142	16.797	8.005	25.687
Rate female over 16	6.644	-10.507	23.608	23.459	6.337	40.772	33.882	23.731	44.341
Median age	0.380	0.235	0.534	0.525	0.378	0.675	0.330	0.252	0.404
Rate adults over 65	-19.822	-31.900	-8.102	-40.757	-52.583	-28.876	-37.559	-43.666	-31.479
Commuting time (min)	0.144	0.102	0.187	0.158	0.115	0.200	0.076	0.054	0.098
Rate white	18.783	2.530	34.230	18.277	2.532	33.703	-14.846	-25.450	-4.168
Rate black/African-American	19.997	3.223	35.841	18.120	1.989	34.181	-18.430	-29.508	-7.506
Rate Asian	7.332	-8.721	22.915	10.595	-5.045	26.118	-5.758	-16.533	5.061
Rate Hawaiian/Pacific islanders	266.396	173.307	356.908	353.610	265.006	443.079	36.369	-22.193	94.272
Rate others	23.809	6.161	40.938	23.965	6.544	41.354	-20.584	-32.446	-8.613
City population	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000

Table G.48: Regression results for testing H3 to see whether the fragmentation of local information is correlated to community engagement in 14 areas over 20 months. DVs are nested within the time.

	Std. Events/Pop			Std. RSVPs/Pop			Std. RSVPs/Event		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H3 (2) (Frag-Engage) ctl. dprv: 14 Areas ($N = 280$)									
(Intercept)	-133.090	-185.780	-81.461	-178.608	-219.997	-136.692	-46.475	-93.713	0.939
The fragmentation of info (β_{CO})	-3.334	-6.009	-0.859	-1.187	-2.294	-0.038	0.332	-1.030	1.660
Socio-economic inequality (Gini)	-41.828	-59.019	-25.054	-49.032	-62.565	-35.222	-6.403	-21.764	9.504
Socio-economic deprivation (depr. index)	-33.544	-39.224	-27.814	-48.944	-53.322	-44.506	-8.000	-13.048	-3.024
Rate foreign-born	424.151	262.073	588.845	507.016	373.721	637.561	149.658	-1.048	298.022
Rate U.S. citizens	172.433	132.028	215.006	239.500	207.764	270.461	51.484	15.335	87.343
Rate female over 16	23.299	-47.734	95.177	5.597	-55.577	65.951	23.751	-43.836	89.970
Median age	-2.069	-2.360	-1.772	-2.964	-3.165	-2.765	-0.222	-0.457	0.015
Rate adults over 65	123.157	90.535	154.988	172.594	145.944	198.719	-2.333	-31.521	27.320
Commuting time (min)	0.445	0.385	0.508	0.560	0.520	0.600	0.134	0.086	0.183
Rate white	16.351	-0.785	33.253	27.773	13.404	42.009	-9.587	-25.502	6.365
Rate black/African-American	8.257	-13.966	30.558	17.835	-1.020	36.683	-14.121	-34.862	6.988
Rate Asian	36.825	24.336	48.974	60.692	51.654	69.810	5.128	-5.306	15.361
Rate others	92.909	76.613	109.235	131.015	119.237	142.892	-0.413	-14.330	13.098
City population	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table G.49: Regression results for testing H3 to see whether the fragmentation of local information is correlated to community engagement in 14 areas over 20 months. DVs are nested within the time. Socio-economic inequality/deprivation is controlled.

	Std. Events/Pop			Std. RSVPs/Pop			Std. RSVPs/Event		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H3 (1) (Frag-Engage): 28 Areas ($N = 84$)									
(Intercept)	-11.029	-33.951	13.544	-17.660	-41.153	6.071	-10.242	-25.256	5.123
The fragmentation of info (β_{CO})	1.207	-3.483	5.866	2.774	-1.962	7.685	-0.635	-3.329	2.008
Rate foreign-born	165.436	107.764	223.746	154.039	95.855	212.253	0.361	-34.006	34.543
Rate U.S. citizens	1.006	-8.211	10.290	0.987	-8.570	10.467	2.355	-3.251	7.877
Rate female over 16	16.820	-6.506	40.078	25.755	1.224	51.394	16.653	3.035	30.752
Median age	0.329	0.132	0.527	0.340	0.137	0.537	0.032	-0.079	0.145
Rate adults over 65	-15.757	-29.703	-1.883	-21.659	-36.038	-7.442	-12.737	-20.810	-4.814
Commuting time (min)	0.166	0.096	0.237	0.137	0.065	0.209	0.085	0.044	0.126
Rate white	-14.923	-32.279	2.041	-12.758	-29.859	4.167	-1.579	-13.673	11.043
Rate black/African-American	-13.923	-30.947	2.969	-12.546	-29.496	4.377	-3.601	-15.771	8.890
Rate Asian	-20.540	-38.657	-2.943	-15.466	-33.213	2.547	8.605	-3.969	21.492
Rate Hawaiian/Pacific islanders	30.717	-73.503	138.060	20.836	-85.774	127.752	-51.469	-119.011	16.753
Rate others	-10.089	-28.643	8.045	-8.182	-26.214	9.513	-9.294	-21.925	3.964
City population	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000

Table G.50: Regression results for testing H3 to see whether the fragmentation of local information is correlated to community engagement in 28 areas over 3 months. DVs are nested within the time.

	Std. Events/Pop			Std. RSVPs/Pop			Std. RSVPs/Event		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H3 (Frag-Engage) ctl. dprv: 28 Areas ($N = 84$)									
(Intercept)	23.972	8.193	39.576	18.876	3.111	34.681	-12.804	-24.428	-1.286
The fragmentation of info (β_{CO})	2.050	-1.875	5.993	5.204	1.212	9.179	2.343	-0.351	5.080
Socio-economic inequality (Gini)	-6.326	-15.209	2.373	0.724	-7.785	9.402	11.231	5.108	17.316
Socio-economic deprivation (depr. index)	-7.986	-10.409	-5.518	-9.812	-12.244	-7.351	-3.355	-5.046	-1.653
Rate foreign-born	85.356	38.402	131.917	64.747	19.273	111.261	-12.946	-45.847	19.426
Rate U.S. citizens	-5.577	-12.743	1.655	-5.206	-12.188	1.835	3.336	-2.178	8.482
Rate female over 16	-19.535	-44.606	6.571	-26.567	-52.372	0.365	-12.332	-30.343	6.264
Median age	-0.070	-0.243	0.101	-0.083	-0.268	0.095	-0.018	-0.145	0.108
Rate adults over 65	24.761	10.958	39.101	20.930	6.696	35.331	-8.159	-18.334	1.659
Commuting time (min)	0.130	0.076	0.184	0.108	0.052	0.165	0.099	0.060	0.140
Rate white	-14.763	-23.689	-5.141	-10.581	-20.036	-1.043	8.029	1.191	15.057
Rate black/African-American	-10.821	-19.848	-1.184	-6.796	-16.607	2.760	7.107	0.207	14.282
Rate Asian	-23.843	-34.237	-12.929	-18.184	-29.244	-7.540	15.678	7.893	23.562
Rate others	-7.784	-17.888	2.605	-2.865	-13.453	7.477	2.081	-5.617	9.945
City population	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000

Table G.51: Regression results for testing H3 to see whether the fragmentation of local information is correlated to community engagement in 28 areas over 3 months. DVs are nested within the time. Socio-economic inequality/deprivation is controlled.

	Std. Events/Pop			Std. RSVPs/Pop			Std. RSVPs/Event		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H3 (1) (Frag-Engage): 14 Areas ($N = 42$)									
(Intercept)	-30.300	-76.927	15.912	-56.058	-103.797	-6.422	-30.972	-63.528	0.905
The fragmentation of info (β_{CO})	8.993	2.155	15.954	10.629	2.867	18.426	-0.245	-4.677	4.155
Rate foreign-born	231.336	69.011	403.727	217.523	41.027	392.915	98.298	-15.934	210.704
Rate U.S. citizens	-2.612	-39.698	36.345	7.361	-33.804	46.106	14.355	-11.275	39.666
Rate female over 16	-5.611	-42.913	32.100	31.852	-7.229	71.783	28.090	3.441	53.696
Median age	0.422	0.066	0.774	0.511	0.113	0.903	0.250	0.020	0.480
Rate adults over 65	-18.033	-45.138	9.561	-41.042	-69.796	-11.623	-32.892	-50.879	-14.922
Commuting time (min)	0.236	0.129	0.342	0.190	0.072	0.309	0.076	0.006	0.145
Rate white	7.353	-12.325	26.866	5.539	-14.041	24.917	-3.380	-21.457	14.495
Rate black/African-American	9.865	-10.351	30.237	4.855	-15.504	25.061	-6.895	-25.670	11.637
Rate Asian	-8.125	-28.632	12.186	-3.464	-23.535	16.526	5.303	-13.070	23.215
Rate Hawaiian/Pacific islanders	237.917	90.780	388.738	302.942	139.662	464.799	116.251	1.789	230.804
Rate others	19.997	-4.034	43.775	14.417	-10.201	39.099	-11.396	-32.018	9.478
City population	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000

Table G.52: Regression results for testing H3 to see whether the fragmentation of local information is correlated to community engagement in 14 areas over 3 months. DVs are nested within the time.

	Std. Events/Pop			Std. RSVPs/Pop			Std. RSVPs/Event		
H3 (Frag-Engage) ctl. dprv: 14 Areas ($N = 42$)									
(Intercept)	-161.484	-211.748	-107.812	-137.406	-202.415	-65.360	-2.760	-71.757	64.075
The fragmentation of info (β_{CO})	-1.704	-3.644	0.402	0.444	-2.774	4.158	-2.346	-7.007	2.382
Socio-economic inequality (Gini)	-51.271	-67.675	-34.318	-38.410	-59.281	-16.565	4.394	-17.863	25.851
Socio-economic deprivation (depr. index)	-42.223	-47.990	-36.080	-42.861	-50.299	-34.867	-11.819	-19.276	-4.243
Rate foreign-born	559.734	394.057	719.692	404.367	188.507	599.479	3.217	-202.240	213.872
Rate U.S. citizens	218.356	172.878	259.517	191.586	123.252	249.934	23.223	-35.213	81.685
Rate female over 16	-0.696	-71.746	68.880	4.067	-72.007	80.248	-21.824	-103.835	61.897
Median age	-2.609	-2.917	-2.247	-2.402	-2.864	-1.824	-0.410	-0.872	0.061
Rate adults over 65	161.409	129.018	193.947	143.874	105.649	181.500	23.200	-14.320	61.141
Commuting time (min)	0.575	0.512	0.632	0.475	0.370	0.565	0.097	-0.000	0.191
Rate white	23.741	7.083	41.015	14.012	-3.409	31.354	-1.584	-18.626	15.499
Rate black/African-American	16.038	-5.754	38.661	6.767	-15.315	28.568	-2.456	-24.944	19.901
Rate Asian	48.097	34.063	61.080	37.912	18.031	55.285	8.069	-7.755	24.131
Rate others	123.452	103.909	140.867	100.010	71.097	124.593	0.269	-22.199	22.759
City population	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table G.53: Regression results for testing H3 to see whether the fragmentation of local information is correlated to community engagement in 14 areas over 3 months. DVs are nested within the time. Socio-economic inequality/deprivation is controlled.

	Rao-Stirling			RS β-weighted			RS α-weighted			Shannon Entropy			HHI of Topics		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H4 (Depr-Cult): 14 Areas ($N = 280$)															
(Intercept)	311.034	257.849	361.991	263.106	207.208	316.604	223.133	169.826	276.356	198.447	141.857	253.667	166.108	108.153	224.440
Socio-economic inequality (Gini)	84.555	67.343	100.809	68.899	51.177	85.795	59.688	42.663	76.670	49.726	31.613	67.178	39.280	20.430	57.841
Socio-economic deprivation (depr. index)	4.853	-0.501	10.187	5.844	0.127	11.273	0.249	-5.443	6.133	2.870	-2.751	8.501	6.596	0.248	12.905
Rate foreign-born	-807.638	-968.518	-640.369	-691.428	-857.733	-515.282	-524.361	-690.403	-357.379	-496.642	-669.852	-320.927	-459.387	-640.575	-275.182
Rate U.S. citizens	-261.594	-301.633	-220.123	-227.452	-269.333	-183.659	-170.830	-213.095	-129.157	-163.875	-207.244	-117.994	-156.456	-203.889	-108.664
Rate female over 16	-234.979	-302.216	-161.965	-182.808	-253.766	-108.551	-200.769	-272.665	-126.406	-151.397	-225.974	-77.106	-89.849	-167.976	-11.668
Median age	1.542	1.262	1.811	1.401	1.106	1.682	0.848	0.557	1.138	0.893	0.588	1.199	1.019	0.661	1.355
Rate adults over 65	-2.816	-34.172	27.825	-7.650	-39.206	24.062	21.146	-12.014	53.361	9.670	-21.648	42.591	-10.482	-45.590	24.843
Commuting time (min)	-0.538	-0.597	-0.479	-0.461	-0.522	-0.397	-0.344	-0.404	-0.286	-0.316	-0.381	-0.249	-0.288	-0.363	-0.215
Rate white	-19.493	-36.154	-3.486	-20.073	-37.069	-3.760	-10.166	-27.983	6.712	-12.845	-29.698	4.568	-15.615	-33.131	2.349
Rate black/African-American	11.562	-10.756	32.198	5.836	-16.463	27.331	13.979	-9.108	35.975	7.663	-14.710	30.545	0.319	-23.072	23.690
Rate Asian	-75.113	-86.379	-63.873	-67.078	-78.679	-55.366	-47.124	-59.605	-35.304	-46.367	-58.708	-34.089	-45.242	-58.442	-31.508
Rate others	-90.304	-105.681	-74.853	-82.219	-98.398	-65.639	-51.751	-68.594	-35.255	-54.707	-71.111	-38.286	-58.448	-76.684	-39.019
City population	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000

Table G.54: Regression results for testing H4 to see whether the socio-economic deprivation/inequality indices are correlated to cultural activity diversity in 14 areas over 20 months. DVs are nested within the time.

	Rao-Stirling			RS β-weighted			RS α-weighted			Shannon Entropy			HHI of Topics		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H4 (Depr-Cult): 28 Areas ($N = 84$)															
(Intercept)	21.997	-5.887	49.782	22.787	-2.166	48.035	32.262	5.609	59.297	29.381	2.587	55.943	23.809	-1.837	48.430
Socio-economic inequality (Gini)	13.371	-1.136	27.758	8.482	-5.271	22.003	3.694	-10.225	17.283	2.792	-11.108	16.357	3.509	-9.066	16.305
Socio-economic deprivation (depr. index)	-4.015	-8.171	0.312	-3.192	-7.039	0.774	-0.474	-4.588	3.540	-1.110	-5.164	2.974	-2.380	-6.059	1.347
Rate foreign-born	-23.950	-110.388	59.977	8.450	-74.776	91.211	20.028	-65.703	105.576	28.509	-54.611	113.111	39.256	-41.465	118.564
Rate U.S. citizens	1.870	-11.254	15.241	1.633	-10.726	14.058	-5.440	-18.429	7.639	-3.751	-16.175	8.772	1.121	-10.620	13.059
Rate female over 16	-45.532	-89.161	-1.144	-47.182	-87.998	-5.804	-52.442	-95.688	-9.845	-50.071	-91.786	-6.299	-48.735	-88.223	-9.882
Median age	0.055	-0.261	0.371	0.127	-0.169	0.427	0.166	-0.139	0.469	0.188	-0.104	0.477	0.200	-0.081	0.479
Rate adults over 65	-7.225	-32.363	17.755	-8.689	-31.866	13.879	-6.860	-31.053	17.066	-7.581	-30.731	16.176	-9.352	-31.525	13.368
Commuting time (min)	-0.075	-0.171	0.026	-0.092	-0.183	-0.000	-0.107	-0.201	-0.010	-0.111	-0.207	-0.014	-0.105	-0.190	-0.019
Rate white	-5.619	-19.778	8.208	-5.427	-18.814	8.471	-5.064	-19.955	9.939	-5.175	-18.974	8.129	-5.439	-18.496	7.897
Rate black/African-American	-1.862	-16.057	11.939	-1.030	-14.677	13.007	0.399	-14.393	15.308	0.264	-13.660	13.547	-0.470	-13.726	12.826
Rate Asian	-6.623	-22.790	9.368	-5.494	-20.891	10.359	-4.638	-21.702	12.175	-4.639	-20.632	10.967	-4.751	-19.872	10.401
Rate others	-14.753	-30.728	0.636	-15.039	-30.485	0.808	-14.943	-31.635	1.559	-15.344	-30.898	0.019	-15.099	-30.058	-0.232
City population	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table G.55: Regression results for testing H4 to see whether the socio-economic deprivation/inequality indices are correlated to cultural activity diversity in 28 areas over 3 months. DVs are nested within the time.

	Rao-Stirling			RS β-weighted			RS α-weighted			Shannon Entropy			HHI of Topics		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H4 (Depr-Cult): 14 Areas ($N = 42$)															
(Intercept)	134.707	44.094	226.767	125.089	32.896	217.526	119.661	35.807	203.099	102.482	17.642	181.591	80.347	-2.433	161.980
Socio-economic inequality (Gini)	46.813	17.400	76.052	38.908	8.098	68.395	26.326	-1.420	53.448	20.299	-7.339	46.431	16.735	-10.143	43.960
Socio-economic deprivation (depr. index)	-14.383	-26.290	-2.219	-13.522	-25.522	-1.597	-12.178	-23.524	-1.065	-11.305	-22.611	0.428	-8.762	-20.359	3.238
Rate foreign-born	-229.171	-499.364	41.865	-215.642	-494.725	67.160	-112.311	-365.083	151.074	-137.891	-380.029	118.435	-145.014	-388.791	109.007
Rate U.S. citizens	-59.161	-133.546	16.296	-58.594	-135.300	17.051	-37.342	-106.232	35.546	-40.029	-106.244	29.130	-46.474	-114.140	23.711
Rate female over 16	-190.049	-297.724	-81.158	-167.357	-276.011	-52.834	-181.846	-282.824	-78.255	-141.035	-243.045	-37.538	-91.988	-197.938	11.381
Median age	0.249	-0.486	0.944	0.240	-0.537	0.981	-0.017	-0.713	0.635	-0.015	-0.717	0.673	0.195	-0.483	0.874
Rate adults over 65	51.590	-3.447	104.644	55.017	-1.467	111.929	71.421	18.792	124.284	67.973	12.698	121.175	49.336	-3.689	102.826
Commuting time (min)	-0.347	-0.513	-0.173	-0.332	-0.501	-0.151	-0.248	-0.400	-0.087	-0.227	-0.375	-0.072	-0.218	-0.372	-0.062
Rate white	-13.981	-31.646	4.696	-13.979	-32.641	4.492	-12.888	-31.156	5.563	-10.689	-28.760	7.920	-9.705	-28.507	8.952
Rate black/African-American	7.031	-15.829	29.931	6.277	-17.531	29.833	8.022	-15.207	31.754	7.475	-15.014	30.575	5.180	-18.295	29.577
Rate Asian	-40.678	-62.334	-18.879	-38.388	-60.442	-15.177	-30.037	-49.970	-9.373	-25.751	-45.350	-5.581	-24.408	-44.117	-4.273
Rate others	-9.128	-40.093	21.731	-7.383	-37.489	24.571	3.673	-25.305	32.197	3.044	-25.490	32.612	-1.173	-30.496	27.731
City population	-0.000	-0.000	0.000	-0.000	-0.000	0.000	-0.000	-0.000	0.000	-0.000	-0.000	0.000	-0.000	-0.000	0.000

Table G.56: Regression results for testing H4 to see whether the socio-economic deprivation/inequality indices are correlated to cultural activity diversity in 14 areas over 3 months. DVs are nested within the time.

	Rao-Stirling			RS β-weighted			RS α-weighted			Shannon Entropy			HHI of Topics		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H5 (Frag-Moderate on Dprv): 14 Areas ($N = 280$)															
(Intercept)	312.719	261.996	365.726	269.201	215.753	322.070	224.117	170.355	278.152	199.184	143.738	255.951	167.895	110.856	225.156
Socio-economic inequality (Gini)	85.214	68.555	102.616	70.889	53.665	88.145	60.433	42.826	77.704	50.371	32.548	68.512	39.507	20.894	57.831
Socio-economic deprivation (depr. index)	17.519	3.725	31.334	19.715	4.802	35.132	1.969	-12.329	16.202	10.766	-4.619	26.170	21.567	2.762	40.794
Rate foreign-born	-806.019	-971.041	-646.909	-703.641	-870.035	-539.350	-526.920	-696.871	-357.536	-495.758	-673.673	-323.642	-457.720	-634.188	-279.731
Rate U.S. citizens	-261.885	-303.465	-221.265	-230.814	-272.685	-187.900	-173.602	-215.399	-131.461	-164.901	-209.392	-119.662	-154.486	-201.701	-108.139
Rate female over 16	-238.197	-311.391	-166.655	-191.492	-264.610	-121.291	-201.718	-276.130	-127.275	-153.234	-228.275	-79.428	-94.029	-166.824	-18.825
Median age	1.553	1.272	1.832	1.400	1.110	1.694	0.882	0.583	1.168	0.911	0.609	1.226	0.987	0.646	1.342
Rate adults over 65	-2.898	-34.116	28.394	-5.416	-36.702	26.510	19.266	-13.063	51.638	8.819	-23.694	40.953	-7.700	-42.232	24.626
Commuting time (min)	-0.538	-0.599	-0.480	-0.465	-0.525	-0.404	-0.343	-0.403	-0.282	-0.313	-0.380	-0.246	-0.287	-0.360	-0.216
Rate white	-19.658	-36.148	-2.927	-18.927	-35.011	-2.348	-10.677	-27.725	6.771	-13.108	-29.989	3.765	-14.839	-32.264	1.539
Rate black/African-American	11.646	-10.108	33.819	7.669	-13.452	29.758	13.643	-8.844	36.685	7.568	-14.788	30.296	1.291	-21.244	22.758
Rate Asian	-75.674	-87.095	-63.857	-66.955	-79.001	-55.018	-48.379	-60.186	-36.059	-47.135	-59.329	-35.348	-44.327	-57.742	-31.171
Rate others	-90.399	-105.873	-74.653	-81.658	-98.022	-65.188	-52.911	-68.727	-36.240	-55.041	-71.795	-38.956	-56.916	-75.660	-38.352
The fragmentation of info (β_{CO})	-0.390	-2.638	1.811	-0.778	-3.347	1.701	1.505	-0.653	3.727	0.543	-2.138	3.047	-1.962	-5.070	1.070
City population	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
Deprivation index:β_{CO}	-13.691	-27.814	0.106	-15.387	-31.145	0.295	-1.208	-15.068	13.016	-8.247	-24.151	7.493	-16.940	-36.509	2.564

Table G.57: Regression results for testing H5 to see whether the fragmentation of local information moderates the effect of socio-economic deprivation on cultural activity diversity in 14 areas over 20 months. DVs are nested within the time.

	Rao-Stirling			RS β-weighted			RS α-weighted			Shannon Entropy			HHI of Topics		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H5 (Frag-Moderate on Dprv): 28 Areas ($N = 84$)															
(Intercept)	29.369	1.447	58.093	28.414	2.070	54.994	38.011	10.283	65.608	34.744	8.153	60.496	26.531	0.677	51.816
Socio-economic inequality (Gini)	5.982	-9.443	21.089	2.741	-12.474	17.721	-1.725	-17.912	14.050	-2.139	-17.379	13.296	-0.027	-13.747	14.136
Socio-economic deprivation (depr. index)	9.567	-20.315	40.406	9.177	-20.094	38.374	7.017	-23.953	37.527	9.216	-20.032	38.284	8.539	-20.221	37.465
Rate foreign-born	-16.921	-104.765	66.022	14.554	-69.906	97.944	23.810	-63.037	109.122	32.051	-52.448	114.910	44.443	-31.561	120.278
Rate U.S. citizens	0.828	-12.424	13.531	1.143	-11.732	13.519	-6.299	-19.179	6.791	-4.634	-17.220	8.218	0.891	-11.304	13.237
Rate female over 16	-30.837	-77.603	12.201	-35.849	-80.128	8.619	-41.325	-86.668	4.266	-40.148	-84.169	4.069	-40.858	-82.067	0.574
Median age	-0.006	-0.328	0.320	0.077	-0.215	0.376	0.125	-0.185	0.432	0.146	-0.156	0.430	0.180	-0.101	0.468
Rate adults over 65	-1.751	-28.019	24.343	-4.300	-27.449	19.501	-3.031	-27.980	21.320	-3.620	-27.022	19.754	-7.512	-30.778	14.781
Commuting time (min)	-0.104	-0.204	-0.003	-0.116	-0.211	-0.018	-0.133	-0.239	-0.029	-0.131	-0.228	-0.035	-0.118	-0.210	-0.026
Rate white	-7.066	-21.272	6.711	-6.549	-20.333	7.205	-6.219	-20.305	7.948	-6.268	-20.127	7.598	-6.310	-19.190	6.957
Rate black/African-American	-4.208	-18.763	10.135	-2.891	-16.999	11.145	-1.446	-15.742	12.721	-1.467	-15.846	12.793	-1.854	-14.991	11.553
Rate Asian	-6.804	-22.770	9.692	-5.568	-21.620	10.383	-4.698	-20.805	11.531	-4.939	-20.928	11.457	-5.122	-20.060	9.963
Rate others	-16.842	-32.563	-1.182	-16.699	-32.365	-1.133	-16.792	-33.335	-0.451	-16.911	-32.593	-0.955	-16.101	-31.010	-1.223
The fragmentation of info (β_{CO})	-7.832	-14.982	-0.723	-6.145	-13.294	0.944	-6.112	-13.055	1.065	-5.447	-12.349	1.505	-3.680	-10.533	3.083
City population	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Deprivation index:β_{CO}	-13.115	-45.830	19.570	-12.154	-43.382	19.332	-6.828	-39.667	26.275	-10.123	-41.532	21.363	-11.024	-42.153	20.532

Table G.58: Regression results for testing H5 to see whether the fragmentation of local information moderates the effect of socio-economic deprivation on cultural activity diversity in 28 areas over 3 months. DVs are nested within the time.

	Rao-Stirling			RS β-weighted			RS α-weighted			Shannon Entropy			HHI of Topics		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H5 (Frag-Moderate on Dprv): 14 Areas ($N = 42$)															
(Intercept)	146.279	56.745	231.136	137.368	48.013	222.717	125.405	44.655	209.567	108.357	26.660	188.421	84.979	2.481	166.939
Socio-economic inequality (Gini)	42.246	12.322	70.940	34.290	4.149	63.161	19.661	-8.777	47.063	14.013	-14.030	41.509	11.034	-17.395	38.812
Socio-economic deprivation (depr. index)	-6.315	-47.346	35.248	-7.021	-46.611	33.064	-7.747	-47.315	32.898	-9.090	-47.125	29.714	-9.332	-48.886	29.666
Rate foreign-born	-243.064	-507.059	29.317	-233.171	-499.852	37.033	-114.389	-365.299	133.182	-142.334	-389.297	106.455	-150.518	-401.662	97.131
Rate U.S. citizens	-52.039	-123.502	21.903	-50.386	-120.967	25.179	-23.641	-93.007	43.792	-28.255	-96.094	41.873	-35.952	-103.518	33.884
Rate female over 16	-189.678	-295.561	-78.365	-169.065	-273.632	-56.058	-177.418	-281.878	-74.399	-135.924	-238.496	-34.040	-86.796	-189.707	17.872
Median age	0.079	-0.668	0.798	0.045	-0.693	0.765	-0.249	-0.942	0.442	-0.216	-0.909	0.445	0.005	-0.702	0.672
Rate adults over 65	65.661	7.160	120.944	71.326	15.045	126.150	88.029	34.302	140.362	82.496	31.227	134.522	62.836	9.465	117.219
Commuting time (min)	-0.387	-0.552	-0.209	-0.375	-0.542	-0.202	-0.283	-0.437	-0.131	-0.262	-0.404	-0.111	-0.248	-0.401	-0.093
Rate white	-14.363	-33.573	4.723	-13.201	-31.924	5.297	-12.191	-30.495	6.655	-10.369	-29.017	8.632	-8.924	-28.016	9.335
Rate black/African-American	6.599	-17.355	30.366	7.193	-16.093	30.331	8.188	-15.103	31.932	7.316	-16.300	31.789	5.427	-19.269	28.909
Rate Asian	-37.846	-59.075	-15.987	-33.891	-55.557	-11.730	-24.449	-44.729	-3.840	-20.874	-40.406	-0.682	-19.516	-39.484	0.881
Rate others	-8.665	-38.669	21.556	-5.288	-35.695	25.636	7.127	-20.653	36.740	5.480	-21.842	34.013	1.553	-26.890	30.174
The fragmentation of info (β_{CO})	-12.963	-24.231	-1.964	-14.332	-25.206	-3.377	-14.112	-24.819	-3.282	-13.202	-23.963	-3.261	-11.677	-22.662	-1.021
City population	-0.000	-0.000	0.000	-0.000	-0.000	0.000	-0.000	-0.000	0.000	-0.000	-0.000	0.000	-0.000	-0.000	0.000
Deprivation index:β_{CO}	-13.948	-62.090	32.217	-12.998	-58.661	32.540	-10.815	-56.514	35.497	-7.699	-52.426	36.554	-4.145	-49.523	40.883

Table G.59: Regression results for testing H5 to see whether the fragmentation of local information moderates the effect of socio-economic deprivation on cultural activity diversity in 14 areas over 3 months. DVs are nested within the time.

	Rao-Stirling			RS β-weighted			RS α-weighted			Shannon Entropy			HHI		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H5 (Frag-Moderate on Gini): 14 Areas ($N = 280$)															
(Intercept)	287.994	232.527	346.346	249.453	191.458	310.306	243.722	185.348	298.584	210.858	153.336	266.275	165.684	103.889	227.441
Socio-economic inequality (Gini)	126.469	88.555	163.160	96.715	56.269	137.113	25.098	-14.683	63.525	27.247	-14.207	69.359	36.935	-11.373	85.143
Socio-economic deprivation (depr. index)	5.143	-0.421	10.603	5.909	-0.013	11.760	0.649	-4.840	6.048	3.161	-2.372	8.814	5.927	-0.116	12.215
Rate foreign-born	-791.912	-963.330	-625.904	-685.242	-862.767	-510.098	-541.082	-703.492	-373.254	-504.664	-669.304	-336.798	-456.462	-634.623	-280.506
Rate U.S. citizens	-258.858	-302.091	-217.324	-226.450	-270.697	-182.421	-176.249	-218.112	-133.718	-166.694	-207.834	-122.952	-153.345	-199.683	-106.229
Rate female over 16	-232.045	-307.244	-160.785	-182.287	-260.992	-106.178	-204.001	-273.894	-129.635	-152.195	-222.427	-79.459	-89.344	-164.583	-14.060
Median age	1.560	1.273	1.840	1.408	1.104	1.711	0.871	0.590	1.153	0.908	0.606	1.207	0.974	0.644	1.327
Rate adults over 65	-4.915	-35.940	27.710	-8.356	-41.740	25.986	20.850	-10.436	51.171	9.144	-22.833	41.513	-7.889	-41.474	26.312
Commuting time (min)	-0.536	-0.597	-0.476	-0.461	-0.526	-0.398	-0.346	-0.408	-0.284	-0.317	-0.378	-0.253	-0.287	-0.359	-0.217
Rate white	-20.940	-37.350	-3.376	-20.699	-38.087	-2.457	-9.498	-26.190	6.595	-12.573	-28.962	4.457	-14.878	-32.148	2.477
Rate black/African-American	9.925	-11.593	32.955	5.211	-17.860	29.418	14.980	-7.600	36.269	8.032	-14.015	30.391	0.875	-21.815	23.600
Rate Asian	-76.235	-87.778	-64.202	-67.611	-80.171	-54.837	-47.493	-58.866	-35.957	-46.594	-58.480	-34.042	-43.755	-57.312	-30.638
Rate others	-91.084	-106.765	-75.156	-82.603	-99.155	-65.156	-52.481	-67.887	-36.884	-55.239	-71.700	-38.278	-56.811	-75.892	-38.656
The fragmentation of info (β_{CO})	21.635	4.446	38.585	13.809	-4.803	32.374	-16.693	-34.935	0.601	-10.906	-29.676	8.952	-2.184	-24.930	19.767
City population	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
gini:beta_co_raw	-45.550	-80.670	-9.286	-29.857	-68.973	9.094	38.405	2.135	76.892	24.590	-16.328	64.397	1.614	-44.589	48.979

Table G.60: Regression results for testing H5 to see whether the fragmentation of local information moderates the effect of socio-economic inequality on cultural activity diversity in 14 areas over 20 months. DVs are nested within the time.

	Rao-Stirling			RS β-weighted			RS α-weighted			Shannon Entropy			HHI		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H5 (Frag-Moderate on Gini): 28 Areas ($N = 84$)															
(Intercept)	18.148	-37.824	74.576	14.817	-44.028	73.291	26.014	-32.862	86.510	18.191	-37.971	76.162	9.961	-46.806	65.612
Socio-economic inequality (Gini)	31.149	-76.931	138.512	34.240	-83.203	150.945	23.467	-89.807	135.376	33.411	-78.781	143.153	37.078	-71.532	149.134
Socio-economic deprivation (depr. index)	-2.412	-6.614	1.863	-1.985	-6.201	2.176	0.886	-3.719	5.420	0.049	-3.909	4.209	-1.523	-5.533	2.407
Rate foreign-born	-19.255	-106.069	66.402	11.942	-73.438	95.115	25.278	-61.231	114.582	32.034	-46.698	111.102	43.087	-33.128	119.439
Rate U.S. citizens	0.729	-12.525	13.879	0.575	-11.775	13.016	-6.342	-19.919	6.933	-4.890	-17.391	7.561	0.579	-10.786	11.969
Rate female over 16	-30.687	-76.293	15.180	-36.016	-80.885	7.600	-40.194	-86.473	5.366	-39.241	-83.879	4.249	-41.216	-82.396	0.014
Median age	-0.016	-0.320	0.293	0.074	-0.220	0.373	0.125	-0.184	0.448	0.149	-0.148	0.444	0.169	-0.119	0.462
Rate adults over 65	-1.002	-25.932	23.247	-3.732	-27.299	19.610	-2.821	-28.244	22.364	-3.629	-27.621	20.059	-6.487	-29.639	16.198
Commuting time (min)	-0.110	-0.211	-0.007	-0.120	-0.221	-0.023	-0.136	-0.242	-0.035	-0.137	-0.236	-0.039	-0.122	-0.213	-0.031
Rate white	-6.968	-21.234	6.839	-6.823	-20.292	6.991	-6.297	-20.589	7.665	-6.440	-20.286	7.617	-5.989	-18.713	7.489
Rate black/African-American	-4.049	-18.566	9.787	-3.046	-16.723	11.204	-1.523	-15.889	13.047	-1.581	-15.849	12.661	-1.445	-14.294	12.163
Rate Asian	-6.247	-22.731	10.243	-5.687	-21.218	10.092	-4.541	-20.962	11.540	-4.845	-20.493	11.360	-4.453	-18.892	10.268
Rate others	-17.227	-32.973	-0.993	-17.410	-33.046	-1.745	-17.032	-33.358	-0.714	-17.397	-32.627	-1.337	-16.183	-30.619	-0.885
The fragmentation of info (β_{CO})	5.126	-50.475	61.071	9.996	-50.775	70.250	7.014	-52.277	64.898	13.134	-45.386	69.677	15.324	-40.917	72.673
City population	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
gini:beta_co_raw	-28.332	-146.590	91.765	-34.836	-163.302	95.498	-28.519	-152.817	97.462	-40.043	-160.659	83.379	-41.351	-163.139	78.160

Table G.61: Regression results for testing H5 to see whether the fragmentation of local information moderates the effect of socio-economic inequality on cultural activity diversity in 28 areas over 3 months. DVs are nested within the time.

	Rao-Stirling			RS β-weighted			RS α-weighted			Shannon Entropy			HHI		
	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%	Est.	CI 5%	CI 95%
H5 (Frag-Moderate on Gini): 14 Areas ($N = 42$)															
(Intercept)	139.215	37.516	237.671	125.948	21.090	229.577	114.984	18.365	211.970	95.042	1.407	192.519	77.210	-22.031	173.665
Socio-economic inequality (Gini)	51.320	-52.569	154.275	43.519	-62.989	148.073	37.199	-65.844	141.464	27.676	-75.159	128.622	23.434	-76.328	120.579
Socio-economic deprivation (depr. index)	-17.876	-30.263	-5.638	-18.045	-30.610	-5.977	-16.743	-28.157	-4.683	-15.192	-26.596	-3.861	-12.843	-24.237	-1.671
Rate foreign-born	-242.226	-512.538	28.506	-217.309	-491.540	58.152	-109.518	-364.810	153.480	-125.336	-372.919	132.377	-144.759	-398.463	116.576
Rate U.S. citizens	-53.196	-124.834	19.540	-46.845	-122.096	28.204	-24.077	-95.774	46.543	-25.088	-94.499	44.927	-34.940	-105.015	35.045
Rate female over 16	-184.997	-289.956	-75.442	-161.773	-271.986	-51.893	-173.717	-276.862	-66.685	-127.006	-228.802	-25.161	-84.440	-186.069	19.146
Median age	0.113	-0.618	0.838	0.053	-0.670	0.774	-0.236	-0.924	0.466	-0.201	-0.867	0.455	0.009	-0.659	0.675
Rate adults over 65	62.503	6.237	119.135	68.650	12.380	123.000	86.662	31.484	138.527	79.311	28.427	130.295	62.233	12.000	112.054
Commuting time (min)	-0.388	-0.554	-0.213	-0.372	-0.541	-0.198	-0.283	-0.434	-0.125	-0.257	-0.408	-0.102	-0.249	-0.404	-0.090
Rate white	-14.659	-34.309	5.165	-14.109	-32.657	4.338	-12.476	-31.257	6.609	-11.758	-30.414	6.466	-9.313	-27.590	9.185
Rate black/African-American	6.091	-18.589	30.582	5.732	-17.301	29.184	7.652	-16.836	32.254	5.247	-18.662	28.983	4.883	-18.460	28.223
Rate Asian	-37.956	-60.977	-14.466	-33.834	-54.981	-12.051	-24.626	-44.836	-3.845	-21.397	-42.039	-1.266	-19.614	-40.619	1.265
Rate others	-9.621	-40.921	21.672	-5.536	-35.249	24.796	6.616	-22.452	35.885	4.586	-23.835	33.515	1.474	-26.570	30.675
The fragmentation of info (β_{CO})	-6.682	-58.254	44.516	-7.616	-60.553	44.877	-3.773	-57.580	48.075	-4.444	-56.145	46.242	-4.780	-56.373	45.746
City population	-0.000	-0.000	0.000	0.000	-0.000	0.000	-0.000	-0.000	0.000	-0.000	-0.000	0.000	-0.000	-0.000	0.000
gini:beta_co_raw	-10.547	-121.146	99.310	-12.136	-125.787	100.397	-20.221	-133.600	94.580	-17.400	-128.027	93.945	-14.469	-120.671	94.990

Table G.62: Regression results for testing H5 to see whether the fragmentation of local information moderates the effect of socio-economic inequality on cultural activity diversity in 14 areas over 3 months. DVs are nested within the time.

Bibliography

Toril Aalberg, Arild Blekesaune, and Eiri Elvestad. Media choice and informed democracy: Toward increasing news consumption gaps in europe? *The International Journal of Press/Politics*, 18(3):281–303, 2013.

Maria Abascal and Delia Baldassarri. Love thy neighbor? ethnoracial diversity and trust reexamined. *American Journal of Sociology*, 121(3):722–782, 2015a.

Maria Abascal and Delia Baldassarri. Love thy neighbor? ethnoracial diversity and trust reexamined. *American Journal of Sociology*, 121(3):722–782, 2015b. doi: 10.1086/683144.

Norazila Abd Aziz, Mohamad Fitri S, Rethinasamy Soubakeavathi, et al. Constructing sustainable digital learning environments for remote rural children of sarawak. *The Journal of Community Informatics*, 11(1), 2015.

Alberto Alesina and Eliana La Ferrara. The determinants of trust. Working Paper 7621, National Bureau of Economic Research, March 2000. URL `http://www.nber.org/papers/w7621`.

Alberto Alesina, Reza Baqir, and Caroline Hoxby. Political jurisdictions in heterogeneous communities. *Journal of political Economy*, 112(2):348–396, 2004.

Rajkumar Arun, Venkatasubramaniyan Suresh, CE Veni Madhavan, and MN Narasimha Murthy. On finding the natural number of topics with latent dirichlet allocation: Some observations. In *Pacific-Asia conference on knowledge discovery and data mining*, pages 391–402. Springer, 2010.

Hurriyet Babacan. Challenges of inclusion: cultural diversity, citizenship and engagement. In *Proceedings of International Conference on Engaging Communities*, pages 1–18. Queensland Department of Main Roads, August 2005.

Lars Backstrom, Eric Sun, and Cameron Marlow. Find me if you can: improving geographical prediction with social and spatial proximity. *Proceedings of the 19th International Conference on World Wide Web*, pages 61–70, 2010.

Christopher A Bail. The cultural environment: Measuring culture with big data. *Theory and Society*, 43(3-4):465–482, 2014.

Christopher A Bail, Taylor W Brown, and Andreas Wimmer. Prestige, proximity, and prejudice: How google search terms diffuse across the world. *American Journal of Sociology*, 124(5):1496–1548, 2019.

Sandra J Ball-Rokeach, Yong-Chan Kim, and Sorin Matei. Storytelling neighborhood: Paths to belonging in diverse urban environments. *Communication Research*, 28(4):392–428, 2001.

Sara Baron and Alexia Strout-Dapaz. Communicating with and empowering international students with a library skills set. *Reference Services Review*, 29(4): 314–326, 2001.

Roland Benabou. Heterogeneity, stratification, and growth: macroeconomic implications of community structure and school finance. *The American Economic Review*, pages 584–609, 1996.

John Carlo Bertot, Brian Real, and Paul T Jaeger. Public libraries building digital inclusive communities: Data and findings from the 2013 digital inclusion survey. *The Library Quarterly*, 86(3):270–289, 2016.

Alessandro Bessi and Emilio Ferrara. Social bots distort the 2016 us presidential election online discussion. *First Monday*, 21(11-7), 2016.

David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.

Joshua Blumenstock, Gabriel Cadamuro, and Robert On. Predicting poverty and wealth from mobile phone metadata. *Science*, 350(6264):1073–1076, 2015.

Perien J Boer. Community meeting the namibian education technology policy with olpc's xo laptops: is it a viable approach? *The Journal of Community Informatics*, 11(1), 2015.

Ann Bonner and Annemaree Lloyd. What information counts at the moment of practice? information practices of renal nurses. *Journal of Advanced Nursing*, 67 (6):1213–1221, 2011.

Karen Bordonaro. Language learning in the library: An exploratory study of esl students. *The Journal of academic librarianship*, 32(5):518–526, 2006.

Christine L Borgman. *From Gutenberg to the global information infrastructure: access to information in the networked world*. Mit Press, 2003.

Geoffrey C Bowker and Susan Leigh Star. Building information infrastructures for social worldsthe role of classifications and standards. In *Community computing and support systems*, pages 231–248. Springer, 1998.

John Brehm and Wendy Rahn. Individual-level evidence for the causes and consequences of social capital. *American journal of political science*, pages 999–1023, 1997.

Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

Leo Breiman. *Classification and regression trees*. Routledge, 2017.

Kerry Brown and Robyn Keast. Citizen-government engagement: community connection through networked arrangements. *Asian Journal of Public Administration*, 25(1):107–131, 2003.

Gary Burnett and Paul T Jaeger. Small worlds, lifeworlds, and information: The ramifications of the information behaviour of social groups in public policy and the public sphere. *Information Research: An International Electronic Journal*, 13 (2), 2008.

Gary Burnett, Paul T Jaeger, and Kim M Thompson. Normative behavior and information: The social aspects of information access. *Library & Information Science Research*, 30(1):56–66, 2008.

Brian S Butler. Membership size, communication activity, and sustainability: A resource-based model of online social structures. *Information systems research*, 12(4):346–362, 2001.

Juan Cao, Tian Xia, Jintao Li, Yongdong Zhang, and Sheng Tang. A density-based method for adaptive lda model selection. *Neurocomputing*, 72(7-9):1775–1781, 2009.

John M Carroll. *The neighborhood in the Internet: Design research projects in community informatics*. Routledge, 2014.

John M Carroll, Michael Horning, Blaine Hoffman, Craig Ganoe, Harold Robinson, and Mary Beth Rosson. Community network 2.0: visions, participation, and engagement in new information infrastructures. *International Symposium on End User Development*, pages 270–275, 2011.

Elfreda A Chatman. Framing social life in theory and research. *The New Review of Information Behaviour Research*, 1(December):3–17, 2000.

Jianfei Chen, Kaiwei Li, Jun Zhu, and Wenguang Chen. Warplda: a cache efficient o (1) algorithm for latent dirichlet allocation. *Proceedings of the VLDB Endowment*, 9(10):744–755, 2016.

Martin L Cody. Bird diversity components within and between habitats in australia. In Robert E Ricklefs and Dolph Schluter, editors, *Species diversity in ecological communities*, pages 147–158. University of Chicago Press, 1993.

Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

Dora L Costa and Matthew E Kahn. Civic engagement and community heterogeneity: An economist's perspective. *Perspectives on Politics*, 1(1):103–111, 2003.

Justin Cranshaw, Raz Schwartz, Jason I Hong, and Norman Sadeh. The livehoods project: Utilizing social media to understand the dynamics of a city. *International AAAI Conference on Web and Social Media*, 2012.

Mary J Culnan. The dimensions of perceived accessibility to information: Implications for the delivery of information systems and services. *Journal of the Association for Information Science and Technology*, 36(5):302–308, 1985.

Nina Czernich, Oliver Falck, Tobias Kretschmer, and Ludger Woessmann. Broadband infrastructure and economic growth. *The Economic Journal*, 121(552):505–532, 2011.

Yves-Alexandre de Montjoye, Arkadiusz Stopczynski, Erez Shmueli, Alex Pentland, and Sune Lehmann. The strength of the strongest ties in collaborative problem solving. *Scientific reports*, 4:5277, 2014.

Adriana De Souza e Silva and Daniel M Sutko. Theorizing locative technologies through philosophies of the virtual. *Communication Theory*, 21(1):23–42, 2011.

Romain Deveaud, Eric SanJuan, and Patrice Bellot. Accurate and effective latent concept modeling for ad hoc information retrieval. *Document numérique*, 17(1): 61–84, 2014.

Peter Thisted Dinesen and Kim Mannemar Sønderskov. Ethnic diversity and social trust: A critical review of the literature and suggestions for a reserach agenda. *The Oxford handbook of social and political trust*, page 175, 2018.

Paul Dourish. *The Stuff of Bits: An Essay on the Materialities of Information*. MIT Press, 2017.

Paul Dourish and Melissa Mazmanian. Media as material: Information representations as material foundations for organizational practice. *Third International Symposium on Process Organization Studies*, page 92, 2011.

Sébastien Dubois. Recognition and renown, the structure of cultural markets: evidence from french poetry. *Journal of Cultural Economics*, 36(1):27–48, 2012.

Michael B Eisenberg. Information literacy: Essential skills for the information age. *DESIDOC Journal of Library & Information Technology*, 28(2):39, 2008.

Yoram Eshet-Alkalai. Digital literacy: A conceptual framework for survival skills in the digital era. *Journal of Educational Multimedia and Hypermedia*, 13(1):93, 2004.

Federal Communication Commissions. FCC Fixed Broadband Deployment Maps. https://broadbandmap.fcc.gov, 2019. (accessed May 3, 2019).

Raya Fidel and Maurice Green. The many faces of accessibility: engineers' perception of information sources. *Information processing & management*, 40(3): 563–581, 2004.

Priya Fielding-Singh. A taste of inequality: Foods symbolic value across the socioeconomic spectrum. *Sociological Science*, 4:424–448, 2017.

Glenn Firebaugh. Empirics of world income inequality. *American Journal of Sociology*, 104(6):1597–1630, 1999.

Karen E Fisher and Charles M Naumer. Information grounds: Theoretical basis and empirical findings on information flow in social settings. In *New Directions in Human Information Behavior*, pages 93–111. Springer, 2006.

Karen E Fisher, Joan C Durrance, and Marian Bouch Hinton. Information grounds and the use of need-based services by immigrants in queens, new york: A context-based, outcome evaluation approach. *Journal of the American Society for Information Science and Technology*, 55(8):754–766, 2004.

Marcus Foth, Laura Forlano, Christine Satchell, Martin Gibbs, and Judith Donath. *From social butterfly to engaged citizen: Urban informatics, social media, ubiquitous computing, and mobile technology to support citizen engagement*. MIT Press, 2011.

Vanessa Frias-Martinez, Victor Soto, Heath Hohwald, and Enrique Frias-Martinez. Characterizing urban landscapes using geolocated tweets. *International Conference on Privacy, Security, Risk and Trust and International Confernece on Social Computing*, pages 239–248, 2012.

David Gabel and Florence Kwan. Accessibility of broadband telecommunication services by various segments of the american population. *Communications policy in transition: The Internet and beyond*, pages 295–320, 2001.

Jonah Gabry and Ben Goodrich. rstanarm: Bayesian applied regression modeling via stan. *R package version 2.18.2*, 2018. URL https://CRAN.R-project.org/package=rstanarm.

Pengjie Gao, Chang Lee, and Dermot Murphy. Financing dies in darkness? the impact of newspaper closures on public finance. *Forthcoming in Journal of Financial Economics (JFE), Retrieved from SSRN: https://ssrn.com/abstract=3175555*, 2018.

Andrew Gelman and John Carlin. Beyond power calculations: Assessing type s (sign) and type m (magnitude) errors. *Perspectives on Psychological Science*, 9 (6):641–651, 2014.

Johanna Gereke, Max Schaub, and Delia Baldassarri. Ethnic diversity, poverty and social trust in germany: Evidence from a behavioral measure of trust. *PLOS ONE*, 13(7):e0199834, 2018. doi: 10.1371/journal.pone.0199834.

Anthony Giddens. *The constitution of society: Outline of the theory of structuration*, volume 349. Univ of California Press, 1986.

Homero Gil de Zúñiga, Nakwon Jung, and Sebastián Valenzuela. Social media use for news and individuals' social capital, civic engagement and political participation. *Journal of Computer-Mediated Communication*, 17(3):319–336, 2012.

Corrado Gini. On the measure of concentration with special reference to income and statistics. *Colorado College Publication, General Series*, 208:73–79, 1936.

Edward L Glaeser, David I Laibson, Jose A Scheinkman, and Christine L Soutter. Measuring trust. *The quarterly journal of economics*, 115(3):811–846, 2000.

Amir Goldberg, Michael T Hannan, and Balázs Kovács. What does it mean to span cultural boundaries? variety and atypicality in cultural consumption. *American Sociological Review*, 81(2):215–241, 2016.

Yuriy Gorodnichenko, Tho Pham, and Oleksandr Talavera. Social media, sentiment and public opinions: Evidence from# brexit and# uselection. Technical report, National Bureau of Economic Research, 2018.

Thomas L Griffiths and Mark Steyvers. Finding scientific topics. *Proceedings of the National academy of Sciences*, 101(suppl 1):5228–5235, 2004.

Tony H Grubesic. The geodemographic correlates of broadband access and availability in the united states. *Telematics and Informatics*, 21(4):335–358, 2004.

Tony H Grubesic. The us national broadband map: Data limitations and implications. *Telecommunications Policy*, 36(2):113–126, 2012.

Jonathan Grudin and John Pruitt. Personas, participatory design and product development: An infrastructure for engagement. In *Proc. PDC*, volume 2002, page 7th, 2002.

Jürgen Habermas. *The structural transformation of the public sphere: An inquiry into a category of bourgeois society.* MIT press, 1991.

Jutta Haider and David Bawden. Conceptions of information poverty in lis: a discourse analysis. *Journal of Documentation*, 63(4):534–557, 2007.

Hui Han, Lee Giles, Hongyuan Zha, Cheng Li, and Kostas Tsioutsiouliklis. Two supervised learning approaches for name disambiguation in author citations. In *Proceedings of the 2004 Joint ACM/IEEE Conference on Digital Libraries, 2004.*, pages 296–305. IEEE, 2004.

Kyungsik Han, Patrick C Shih, Mary Beth Rosson, and John M Carroll. Understanding local community attachment, engagement and social support networks mediated by mobile technology. *Interacting with Computers*, 28(3):220–273, 2014.

Michael T Hannan and Glenn R Carroll. *Dynamics of organizational populations: Density, legitimation, and competition*. Oxford University Press, 1992.

Michael T Hannan and John Freeman. *Organizational ecology*. Harvard University Press, 1993.

Kotaro Hara, Vicki Le, and Jon Froehlich. Combining crowdsourcing and google street view to identify street-level accessibility problems. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 631–640, 2013.

Steve Harrison and Paul Dourish. Re-place-ing space: the roles of place and space in collaborative systems. *Proceedings of the ACM conference on Computer Supported Cooperative Work*, pages 67–76, 1996.

Brian W Head. Community engagement: participation on whose terms? *Australian Journal of Political Science*, 42(3):441–454, 2007.

John F Helliwell and Robert D Putnam. The social context of well-being. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 359(1449):1435, 2004.

Martin Hilbert. Technological information inequality as an incessantly moving target: The redistribution of information and communication capacities between 1986 and 2010. *Journal of the Association for Information Science and Technology*, 65(4):821–835, 2014.

Heather Hill. Disability and accessibility in the library and information science literature: A content analysis. *Library & Information Science Research*, 35(2): 137–142, 2013.

Albert O Hirschman. The paternity of an index. *The American economic review*, 54(5):761–762, 1964.

Blaine Hoffman, Harold Robinson, Keith Han, and John Carroll. Civicinity events: pairing geolocation tools with a community calendar. *Proceedings of the 3rd International Conference on Computing for Geospatial Research and Applications*, page 14, 2012.

Geert Hofsteds. *Culture's consequences*. Beverly Hills, Sage Publications, 1980.

Mohammad Delwar Hossain. Users and uses of internet access points in bangladesh: A case study of community information centers (CICs). *The Journal of Community Informatics*, 11(1), 2015.

Desislava Hristova, Luca M. Aiello, and Daniele Quercia. The new urban success: How culture pays. *Frontiers in Physics*, 6, 2018. doi: 10.3389/fphy.2018.00027.

Yingjie Hu, Chengbin Deng, and Zhou Zhou. A semantic and sentiment analysis on online neighborhood reviews for understanding the perceptions of people toward their living environments. *Annals of the American Association of Geographers*, pages 1–21, 2019.

Yuheng Hu, Shelly D Farnham, and Andrés Monroy-Hernández. Whoo. ly: facilitating information seeking for hyperlocal communities using social media. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3481–3490, 2013.

Lee Humphreys. Mobile social networks and urban public space. *New Media & Society*, 12(5):763–778, 2010.

Paul Jaccard. The distribution of the flora in the alpine zone. 1. *New phytologist*, 11(2):37–50, 1912.

Paul T Jaeger and Gary Burnett. Information access and exchange among small worlds in a democratic society: the role of policy in shaping information behavior in the post-9/11 united states. *The Library Quarterly*, 75(4):464–495, 2005.

Paul T Jaeger and Gary Burnett. *Information worlds: Behavior, technology, and social context in the age of the Internet.* Routledge, 2010.

Isaac L Johnson, Yilun Lin, Toby Jia-Jun Li, Andrew Hall, Aaron Halfaker, Johannes Schöning, and Brent Hecht. Not at home on the range: Peer production and the urban/rural divide. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 13–25, 2016.

Lou Jost. Partitioning diversity into independent alpha and beta components. *Ecology*, 88(10):2427–2439, 2007.

Dean S. Karlan. Social Capital and Group Banking. Working Paper 181, Princeton University, Woodrow Wilson School of Public and International Affairs, Research Program in Development Studies., May 2005. URL `https://ideas.repec.org/p/pri/rpdevs/karlan_d_soccap_grp_bankingpaper.pdf.html`.

Andrea Kavanaugh, Siddarth Krishnan, Manuel Pérez-Quiñones, John Tedesco, Kumbirai Madondo, and Ankit Ahuja. Encouraging civic participation through local news aggregation. *Information Polity*, 19(1, 2):35–56, 2014.

CJ Keylock. Simpson diversity and the shannon–wiener index as special cases of a generalized entropy. *Oikos*, 109(1):203–207, 2005.

Safirotu Khoir, Jia Tina Du, and Andy Koronios. Study of Asian Immigrants' Information Behaviour in South Australia: Preliminary Results. *Proceedings of the 2014 iConference 2014*, 2014.

Jinseok Kim and Jana Diesner. Distortive effects of initial-based name disambiguation on measurements of large-scale coauthorship networks. *Journal of the Association for Information Science and Technology*, 67(6):1446–1461, 2016.

Yong-Chan Kim and Sandra J Ball-Rokeach. Civic engagement from a communication infrastructure perspective. *Communication Theory*, 16(2):173–197, 2006.

Kim-Mai Cutler. Here's A Detailed Breakdown Of Racial And Gender Diversity Data Across U.S. Venture Capital Firms. `https://techcrunch.com/2015/10/06/s23p-racial-gender-diversity-venture/`, 2015. (accessed May 31, 2019).

Rob Kitchin. *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage, 2014.

Patricia Koleff, Kevin J. Gaston, and Jack J. Lennon. Measuring beta diversity for presenceabsence data. *Journal of Animal Ecology*, 72(3):367–382, 2003.

Lee Komito and Jessica Bates. Migrants' information practices and use of social media in ireland: networks and community. *Proceedings of the 2011 iConference*, pages 289–295, 2011.

Matthias Korn and Amy Voida. Creating friction: infrastructuring civic engagement in everyday life. In *Proceedings of The Fifth Decennial Aarhus Conference on Critical Alternatives*, pages 145–156. Aarhus University Press, 2015.

Kenneth L Kraemer, Dale Ganley, and Sanjeev Dewan. Across the digital divide: A cross-country multi-technology analysis of the determinants of it penetration. *Journal of the Association for Information Systems*, 6(12):10, 2005.

Borje Langefors. Discussion of the Article by Bostrom and Heinen: MIS Problems and Failures: A Socio-Technical Perspective. Part I: The Causes. *MIS Quarterly*, 2(2):55–62, 1978.

James Laurence. The effect of ethnic diversity and community disadvantage on social cohesion: A multi-level analysis of social capital and interethnic relations in uk communities. *European Sociological Review*, 27(1):70–89, 2009.

James Laurence, Katharina Schmid, and Miles Hewstone. Ethnic diversity, ethnic threat, and social cohesion:(re)-evaluating the role of perceived out-group threat and prejudice in the relationship between community ethnic diversity and intra-community cohesion. *Journal of Ethnic and Migration Studies*, 45(3):395–418, 2019.

Christopher A Le Dantec, Kari E Watkins, Russ Clark, and Elizabeth Mynatt. Cycle Atlanta and OneBusAway: Driving innovation through the data ecosystems of civic computing. *International Conference on Human-Computer Interaction*, pages 327–338, 2015.

Danielle Hyunsook Lee and Peter Brusilovsky. Improving personalized recommendations using community membership information. *Information Processing & Management*, 53(5):1201–1214, 2017.

Myeong Lee and Brian S Butler. How are information deserts created? A theory of local information landscapes. *Journal of the Association for Information Science and Technology*, 70(2):101–116, 2019.

Myeong Lee, Rachael Dottle, Carlos Espino, Imam Subkhan, Ariel Rokem, and Afra Mashhadi. A tool for estimating and visualizing poverty maps. *NetMob 2017*, 2017.

Jack J Lennon, Patricia Koleff, JJD Greenwood, and Kevin J Gaston. The geographical structure of british bird distributions: diversity, spatial turnover and scale. *Journal of Animal Ecology*, 70(6):966–979, 2001.

Paul M Leonardi and Stephen R Barley. Materiality and change: Challenges to building better theory about technology and organizing. *Information and Organization*, 18(3):159–176, 2008.

Paul M Leonardi, Bonnie A Nardi, and Jannis Kallinikos. *Materiality and organizing: Social interaction in a technological world*. Oxford University Press on Demand, 2012.

Jessica Lingel. Information tactics of immigrants in urban environments. *Information Research*, 16(4), 2011.

Mengxiong Liu and Bernice Redfern. Information-seeking behavior of multicultural students: A case study at san jose state university. *College & Research Libraries*, 58(4):348–354, 1997.

Annemaree Lloyd. Information literacy: different contexts, different concepts, different truths? *Journal of Librarianship and Information Science*, 37(2):82–88, 2005.

Annemaree Lloyd. Information literacy landscapes: an emerging picture. *Journal of Documentation*, 62(5):570–583, 2006.

Annemaree Lloyd. *Information literacy landscapes: Information literacy in education, workplace and everyday contexts*. Elsevier, 2010.

Annemaree Lloyd. Following the red thread of information in information literacy research: Recovering local knowledge through interview to the double. *Library & Information Science Research*, 36(2):99–105, 2014.

Annemaree Lloyd. Researching fractured (information) landscapes: Implications for library and information science researchers undertaking research with refugees and forced migration studies. *Journal of Documentation*, 73(1):35–47, 2017.

Claudia López and Rosta Farzan. Lend me sugar, i am your neighbor!: a content analysis of online forums for local communities. *Proceedings of the 7th International Conference on Communities and Technologies*, pages 59–67, 2015.

Claudia López, Brian Butler, and Peter Brusilovsky. Does anything ever happen around here? assessing the online information landscape for local events. *Journal of Urban Technology*, 21(4):95–123, 2014.

Meethu Malu and Leah Findlater. Personalized, wearable control of a head-mounted display for users with upper body motor impairments. *Proceedings of the 33rd ACM Conference on Human Factors in Computing Systems*, pages 221–230, 2015.

Christopher D Manning, Christopher D Manning, and Hinrich Schütze. *Foundations of statistical natural language processing*. MIT press, 1999.

Mikhil Masli, Landon Bouma, Andrew Owen, and Loren Terveen. Geowiki+ route analysis= improved transportation planning. *Proceedings of the 2013 Conference on Computer Supported Cooperative Work Companion*, pages 213–218, 2013.

Josef A Mazanec, John C Crotts, Dogan Gursoy, and Lu Lu. Homogeneity versus heterogeneity of cultural values: An item-response theoretical approach applying hofstede's cultural dimensions in a single nation. *Tourism Management*, 48:299–304, 2015.

Grant McKenzie, Zheng Liu, Yingjie Hu, and Myeong Lee. Identifying urban neighborhood names through user-contributed online property listings. *ISPRS International Journal of Geo-Information*, 7(10):388, 2018.

Tom van der Meer and Jochem Tolsma. Ethnic diversity and its effects on social cohesion. *Annual Review of Sociology*, 40(1):459–478, 2014. doi: 10.1146/annurev-soc-071913-043309.

Bruce Mitchell and Juan Franco. Holc redlining maps: The persistent structure of segregation and economic inequality. *NCRC Research*, 2018. URL `https://ncrc.org/holc/`.

Ganesan Muthiah. Assessment of mobile voice agricultural messages given to farmers of cauvery delta zone of tamil nadu, india. *The Journal of Community Informatics*, 11(1), 2015.

Tommaso Nannicini, Andrea Stella, Guido Tabellini, and Ugo Troiano. Social capital and political accountability. *American Economic Journal: Economic Policy*, 5(2): 222–50, 2013.

Murzintcev Nikita. ldatuning: Tuning of the latent dirichlet allocation models parameters. *R package version 1.0.0*, 2019. URL `https://CRAN.R-project.org/package=ldatuning`.

Pippa Norris. *Digital divide: Civic engagement, information poverty, and the Internet worldwide*. Cambridge University Press, 2001.

Anastasios Noulas, Salvatore Scellato, Renaud Lambiotte, Massimiliano Pontil, and Cecilia Mascolo. A tale of many cities: universal patterns in human urban mobility. *PLOS ONE*, 7(5):e37027, 2012.

Chi Young Oh, Brian S Butler, and Myeong Lee. Information behavior of international students settling in an unfamiliar geo-spatial environment. *Proceedings of the American Society for Information Science and Technology*, 51(1):1–11, 2014.

Gianmarco Ottaviano and Giovanni Peri. The economic value of cultural diversity: evidence from US cities. *Journal of Economic geography*, 6(1):9–44, 2006.

Minsu Park, Ingmar Weber, Mor Naaman, and Sarah Vieweg. Understanding musical diversity via online social media. In *Ninth International AAAI Conference on Web and Social Media*, 2015.

Minsu Park, Jaram Park, Young Min Baek, and Michael Macy. Cultural values and cross-cultural video consumption on youtube. *PLoS one*, 12(5):e0177865, 2017.

Josh Pasek, Eian More, and Daniel Romer. Realizing the social internet? online social networking meets offline civic engagement. *Journal of Information Technology & Politics*, 6(3-4):197–215, 2009.

Matteo Pedrini, Valentina Bramanti, and Benedetto Cannatelli. The impact of national culture and social capital on corporate social responsibility attitude among immigrants entrepreneurs. *Journal of Management & Governance*, 20(4):759–787, 2016.

Robert K Peet. The measurement of species diversity. *Annual review of ecology and systematics*, 5(1):285–307, 1974.

Richard A Peterson and Roger M Kern. Changing highbrow taste: From snob to omnivore. *American sociological review*, pages 900–907, 1996.

Karen E Pettigrew. Waiting for chiropody: contextual results from an ethnographic study of the information behaviour among attendees at community clinics. *Information Processing & Management*, 35(6):801–817, 1999.

Karen E Pettigrew, Joan C Durrance, and Kenton T Unruh. Facilitating community information seeking using the internet: Findings from three public library–community network systems. *Journal of the American Society for Information Science and Technology*, 53(11):894–903, 2002.

Thomas F Pettigrew. Intergroup contact theory. *Annual review of psychology*, 49 (1):65–85, 1998.

Elizabeth G Pontikes and Michael T Hannan. An ecology of social categories. *Sociological science.*, 1:311–343, 2014.

Robert D Putnam. E pluribus unum: Diversity and community in the twenty-first century the 2006 Johan Skytte Prize Lecture. *Scandinavian Political Studies*, 30 (2):137–174, 2007.

Giovanni Quattrone, Licia Capra, and Pasquale De Meo. There's no such thing as the perfect map: Quantifying bias in spatial crowd-sourcing datasets. *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work*, pages 1021–1032, 2015.

Daniele Quercia, Joao Paulo Pesce, Virgilio Almeida, and Jon Crowcroft. Psychological maps 2.0: a web engagement enterprise starting in london. *Proceedings of the 22nd International Conference on World Wide Web*, pages 1065–1076, 2013.

Lisa Quirke. Information practices in newcomer settlement: a study of Afghan immigrant and refugee youth in Toronto. *Proceedings of the 2012 iConference*, pages 535–537, 2012.

C Radhakrishna Rao. Diversity: its measurement, decomposition apportionment and analysis. *Sankhy : The Indian Journal of Statistics, Series A*, 44(1):1–22, 1982.

Tom Risen. Where Should You Start Your Tech Company?: A new study shows that the best city for tech isnt́ necessarily Silicon Valley. *U.S. News*, 2016. URL `https://www.usnews.com/news/articles/2016-05-11/where-should-you-start-your-tech-company`. (accessed May 3, 2019).

Richard D Routledge. On whittaker's components of diversity. *Ecology*, 58(5): 1120–1127, 1977.

Helen Russell and Kirsty Young. Influences and experiences of using digital devices in laterlife. *The Journal of Community Informatics*, 11(1), 2015.

Reijo Savolainen. Small world and information grounds as contexts of information seeking and sharing. *Library & Information Science Research*, 31(1):38–45, 2009.

Merlin Schaeffer. Can competing diversity indices inform us about why ethnic diversity erodes social cohesion? a test of five diversity indices in germany. *Social Science Research*, 42(3):755–774, 2013. doi: 10.1016/j.ssresearch.2012.12.018.

Herbert Schiller. *Information inequality.* Routledge, 2013.

Douglas Schuler. Engaging academia: Strengthening the link between community and technology. *The Journal of Community Informatics*, 11(2), 2015.

Dmitriy Selivanov and Qing Wang. text2vec: Modern text mining framework for r. *R package version 0.5.1*, 2018. URL `https://CRAN.R-project.org/package=text2vec`.

Shilad W Sen, Heather Ford, David R Musicant, Mark Graham, Oliver SB Keyes, and Brent Hecht. Barriers to the localness of volunteered geographic information. *Proceedings of the 33rd ACM Conference on Human Factors in Computing Systems*, pages 197–206, 2015.

Lauren F Sessions. How offline gatherings affect online communities: when virtual community members meetup. *Information, Communication & Society*, 13(3): 375–395, 2010.

Claude Elwood Shannon. A mathematical theory of communication. *Bell system technical journal*, 27(3):379–423, 1948.

Ben Shneiderman. Universal usability. *Communications of the ACM*, 43(5):84–91, 2000.

Snunith Shoham and Sarah Kaufman. Information needs of north american immigrants to israel. *Journal of Information, Communication and Ethics in Society*, 5(2/3):185–205, 2007.

Carson Sievert and Kenneth Shirley. Ldavis: A method for visualizing and interpreting topics. In *Proceedings of the workshop on interactive language learning, visualization, and interfaces*, pages 63–70, 2014.

Sei-Ching Joanna Sin and Kyung-Sun Kim. International students' everyday life information seeking: The informational value of social networking sites. *Library & Information Science Research*, 35(2):107–116, 2013.

Judith D Singer and John B Willett. *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford university press, 2003.

Matthew David Smith and Jo Hanisch. Impacts of Mobile Phone Use on Poverty Reduction for Non-Commercial Farmers in Rural North Ghana. *The Journal of Community Informatics*, 11(1), 2015.

Chris Smith-Clarke and Licia Capra. Beyond the baseline: Establishing the value in mobile phone based poverty estimates. In *Proceedings of the 25th international conference on world wide web*, pages 425–434. International World Wide Web Conferences Steering Committee, 2016.

Victor Soto, Vanessa Frias-Martinez, Jesus Virseda, and Enrique Frias-Martinez. Prediction of socioeconomic levels using cell phone records. *International Conference on User Modeling, Adaptation, and Personalization*, pages 377–388, 2011.

Andy Stirling. A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society Interface*, 4(15):707–719, 2007.

Dietlind Stolle and Allison Harell. Social capital and ethno-racial diversity: Learning to trust in an immigrant society. *Political Studies*, 61(1):42–66, 2013.

Dietlind Stolle, Stuart Soroka, and Richard Johnston. When does diversity erode trust? neighborhood diversity, interpersonal trust and the mediating effect of social interactions. *Political studies*, 56(1):57–75, 2008.

Alexander Strehl, Joydeep Ghosh, and Raymond Mooney. Impact of similarity measures on web-page clustering. In *Workshop on artificial intelligence for web search (AAAI 2000)*, volume 58, page 64, 2000.

Mirjam Struppek. The social potential of urban screens. *Visual Communication*, 5 (2):173–188, 2006.

Patrick Sturgis, Ian Brunton-Smith, Sanna Read, and Nick Allum. Does ethnic diversity erode trust? putnams hunkering down thesis reconsidered. *British Journal of Political Science*, 41(1):57–82, 2011. doi: 10.1017/S0007123410000281.

Patrick Sturgis, Ian Brunton-Smith, Jouni Kuha, and Jonathan Jackson. Ethnic diversity, segregation and the social cohesion of neighbourhoods in london. *Ethnic and Racial Studies*, 37(8):1286–1309, 2014. doi: 10.1080/01419870.2013.831932.

Henri Tajfel, John C Turner, William G Austin, and Stephen Worchel. An integrative theory of intergroup conflict. *Organizational identity: A reader*, pages 56–65, 1979.

Harsh Taneja and Utsav Mamoria. Measuring media use across platforms: Evolving audience information systems. *International Journal on Media Management*, 14 (2):121–140, 2012.

Alex S Taylor, Siân Lindley, Tim Regan, David Sweeney, Vasillis Vlachokyriakos, Lillie Grainger, and Jessica Lingel. Data-in-place: Thinking through the relations between data and community. *Proceedings of the 33rd ACM Conference on Human Factors in Computing Systems*, pages 2863–2872, 2015.

Kjerstin Thorson and Chris Wells. Curated flows: A framework for mapping media exposure in the digital age. *Communication Theory*, 26(3):309–328, 2015.

Paulina O Tindana, Jerome A Singh, C Shawn Tracy, Ross EG Upshur, Abdallah S Daar, Peter A Singer, Janet Frohlich, and James V Lavery. Grand challenges in global health: community engagement in research in developing countries. *PLOS Medicine*, 4(9):e273, 2007.

Ciaran B Trace. Information creation and the notion of membership. *Journal of Documentation*, 63(1):142–164, 2007.

Ciaran B Trace. Resistance and the underlife: Informal written literacies and their relationship to human information behavior. *Journal of the Association for Information Science and Technology*, 59(10):1540–1554, 2008.

Eric M. Uslaner. *The Oxford Handbook of Social and Political Trust.* Oxford University Press, 2018. ISBN 978-0-19-086781-2.

Alexander JaM Van Deursen and Ellen J Helsper. A nuanced understanding of internet use and non-use among the elderly. *European journal of communication*, 30(2):171–187, 2015.

Alexander JAM van Deursen and Jan AGM van Dijk. The first-level digital divide shifts from inequalities in physical access to inequalities in material access. *new media & society*, 21(2):354–375, 2019.

Jan Van Dijk. Widening information gaps and policies of prevention. *Digital Democracy: Issues of Theory and Practice*, pages 166–183, 2000.

Jan Van Dijk. *The deepening divide: Inequality in the information society*. Sage Publications, 2005.

Alessandro Venerandi, Giovanni Quattrone, Licia Capra, Daniele Quercia, and Diego Saez-Trumper. Measuring urban deprivation from user generated content. *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 254–264, 2015.

Renee E Walker, Christopher R Keane, and Jessica G Burke. Disparities and access to healthy food in the United States: A review of food deserts literature. *Health & place*, 16(5):876–884, 2010.

Gang Wang, Sarita Yardi Schoenebeck, Haitao Zheng, and Ben Y Zhao. "Will Check-in for Badges": Understanding Bias and Misbehavior on Location-Based Social Networks. *International AAAI Conference on Web and Social Media*, pages 417–426, 2016a.

Ping Wang, Myeong Lee, Xu Meng, and Brian Butler. Toward an ecology theory of creativity in it products: A study of mobile device industry. In *Proceedings of the International Conference on Information Systems (ICIS)*, pages 1–20, 2016b.

James G Webster and Thomas B Ksiazek. The dynamics of audience fragmentation: Public attention in an age of digital media. *Journal of communication*, 62(1):39–56, 2012.

Lilian Weng, Jacob Ratkiewicz, Nicola Perra, Bruno Gonçalves, Carlos Castillo, Francesco Bonchi, Rossano Schifanella, Filippo Menczer, and Alessandro Flammini. The role of information diffusion in the evolution of social networks. *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 356–364, 2013.

Robert Harding Whittaker. Vegetation of the siskiyou mountains, oregon and california. *Ecological monographs*, 30(3):279–338, 1960.

Kenneth C Wilbur. A two-sided, empirical model of television advertising and viewing markets. *Marketing science*, 27(3):356–378, 2008.

Kirsty Williamson. Discovered by chance: The role of incidental information acquisition in an ecological model of information use. *Library & information science research*, 20(1):23–40, 1998.

Dejian Yu, Wanru Wang, Shuai Zhang, Wenyu Zhang, and Rongyu Liu. Hybrid self-optimized clustering model based on citation links and textual features to detect research topics. *PloS one*, 12(10):e0187164, 2017.

Liangzhi Yu. Understanding information inequality: Making sense of the literature of the information and digital divides. *Journal of Librarianship and Information Science*, 38(4):229–252, 2006.

J David L Zakus and Catherine L Lysack. Revisiting community participation. *Health policy and planning*, 13(1):1–12, 1998.

Pengyu Zhu. Telecommuting, household commute and location choice. *Urban Studies*, 50(12):2441–2459, 2013.

Robert W Zmud and L Eugene Apple. Measuring technology incorporation/infusion. *Journal of Product Innovation Management*, 9(2):148–155, 1992.